

PROTEINS AND NUCLEIC ACIDS ENCODING SAME

FIELD OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded thereby.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as NOVX, or NOV1-99 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated NOVX nucleic acid molecule encoding a NOVX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2n, wherein n is an integer between 1 and 162. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162) or a complement of said oligonucleotide.

Also included in the invention are substantially purified NOVX polypeptides (SEQ ID NOS:2n, wherein n is an integer between 1 and 162). In certain embodiments, the NOVX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

5 The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide,
10 or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression
15 of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex
20 between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

Also included in the invention is a method of detecting the presence of a NOVX
25 nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a
30 compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing various disorders or syndromes described below.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

5 For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in
10 gene therapy, and NOVX may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

The invention further includes a method for screening for a modulator of disorders or
15 syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

20 Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a
25 NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the
30 control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the

amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identify the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX

polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

5

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1a	CG56592-01	1	2	Claudin 6-like
1b	CG56586-01	3	4	Claudin-3-like
1c	CG56592-03	5	6	Claudin-6-like
1d	CG56592-02	7	8	Claudin 6-like
2	CG56596-01	9	10	Protein Serine Kinase -like
3a	CG56594-01	11	12	Claudin-19-like
3b	CG56594-02	13	14	Claudin-19-like
3c	CG57576-01	15	16	Claudin-19-like
4a	CG56589-01	17	18	Claudin-6-like
4b	CG56589-01	19	20	Claudin-6-like
4c	CG56589-02	21	22	Claudin-6-like
5a	CG56635-01	23	24	Monocarboxylate transporter (MCT3)-like
5b	CG56635-02	25	26	Monocarboxylate transporter 3-like
5c	CG56635-03	27	28	Monocarboxylate transporter 3-like
5d	CG56635-04	29	30	Monocarboxylate transporter 3-like
5e	CG56635-05	31	32	Monocarboxylate transporter 3-like
6	CG56674-01	33	34	Nitrilase-1-like
7a	CG56613-01	35	36	Cleavage Signal-1 Protein-Like
7b	CG56613-02	37	38	Cleavage Signal-1 Protein-Like
7c	CG56613-03	39	40	Cleavage Signal-1 Protein-Like
7d	174307820	41	42	Cleavage Signal-1 Protein-Like
7e	167474749	323	324	Cleavage Signal-1 Protein-Like
8	153472451	43	44	Matriptase -like
9a	CG56554-01	45	46	Neuropeptide Y/Peptide YY receptor -like
9b	CG56554-02	47	48	Neuropeptide Y/Peptide YY receptor -like
10	CG55964-01	49	50	G-Protein Coupled Receptor-like
11	CG55966-01	51	52	G-Protein Coupled Receptor-like
12	CG56003-01	53	54	Neuromodulin-like
13a	CG56021-01	55	56	G-Protein Coupled Receptor-like
13b	CG56021-02	57	58	G-Protein Coupled Receptor-like
14	CG56023-01	59	60	G-Protein Coupled Receptor-like
15a	CG56065-01	61	62	G-Protein Coupled Receptor-

				like
15b	CG56065-02	63	64	G-Protein Coupled Receptor-like
16a	CG56067-01	65	66	G-Protein Coupled Receptor-like
16b	CG56753-02	67	68	G-Protein Coupled Receptor-like
17a	CG56657-01	69	70	G-Protein Coupled Receptor-like
17b	CG56657-02	71	72	G-Protein Coupled Receptor-like
17c	CG56659-01	73	74	G-Protein Coupled Receptor-like
17d	CG56659_02	75	76	G-Protein Coupled Receptor-like
18a	CG56663-01	77	78	G-Protein Coupled Receptor-like
18b	CG56663-02	79	80	G-Protein Coupled Receptor-like
19a	CG56665-01	81	82	G-Protein Coupled Receptor-like
19b	CG56665-02	83	84	G-Protein Coupled Receptor-like
20	CG56667-01	85	86	G-Protein Coupled Receptor-like
21a	CG56639-01	87	88	Adrenal Secretory Serine Protease-Like
21b	CG56639-02	89	90	Adrenal Secretory Serine Protease-Like
22a	CG56643-01	91	92	Adrenal Secretory Serine Protease-Like
22b	CG56643-02	93	94	Adrenal Secretory Serine Protease-Like
22c	CG56643-03	95	96	Adrenal Secretory Serine Protease-Like
23a	CG56647-02	97	98	Serine Protease DESC1-like
23b	CG56647-03	99	100	Serine Protease DESC1-like
23c	CG56647-01	101	102	Serine Protease DESC1-like
24a	CG56155-01	103	104	Parchorin-like
24b	CG56155-02	105	106	Parchorin-like
25	CG56457-01	107	108	Protein Phosphatase-like
26a	CG56461-01	109	110	GAGE-7-like
26b	CG56461-02	111	112	GAGE-7-like
27a	CG56645-01	113	114	Sodium-Glucose Cotransporter-like
27b	CG56645-02	115	116	Sodium-Glucose Cotransporter-like
27c	191828203	117	118	Sodium-Glucose Cotransporter-like
28	CG56185-01	119	120	MYD-1-like
29a	CG56187-01	121	122	CRAL-TRIO-like
29b	CG56187-03	123	124	CRAL-TRIO-like
29c	CG56189-01	125	126	CRAL-TRIO-like
30	CG56191-01	127	128	Ryudocan-like
31	CG56392-01	129	130	Sulfur-rich Keratin-like
32	CG56686-01	131	132	DNMT1 associated protein-1 (DMAP)
33	CG56688-01	133	134	Notch1-like
34	CG56715-01	135	136	Olfactory Receptor-like
35	CG56718-01	137	138	Olfactory Receptor-like
36a	CG56729-01	139	140	Cadherin 11-like
36b	CG56729-02	141	142	Cadherin 11-like
37	CG56733-01	143	144	Ten-M2-like
38	CG56737-01	145	146	Activin Beta C Chain-like
39a	CG56737-02	147	148	Activin Beta C Chain-like

39b	CG56637-03	149	150	Inhibin Beta E Chain-like
40	CG56097-01	151	152	UDP Glycosyltransferase-like
41a	CG56680-01	153	154	Sodium/Hydrogen Exchanger 4-like
41b	CG56680-02	155	156	Sodium/Hydrogen Exchanger 4-like
42a	CG56682-01	157	158	Kupffer Cell Receptor-like
42b	CG56682-02	159	160	Kupffer Cell Receptor-like
42c	CG56682-03	161	162	Kupffer Cell Receptor-like
42d	CG56682-04	163	164	Kupffer Cell Receptor-like
43	CG56690-01	165	166	P2Y Purinoceptor -like
44	CG56692-01	167	168	G Protein Coupled Receptor-like
45	CG56694-01	169	170	Mas Proto-Oncogene-like
46a	CG56696-01	171	172	Mas Proto-Oncogene-like
46b	CG56696-02	173	174	Mas-Related G Protein-Coupled Receptor-like
46c	CG56702-01	175	176	Mas Proto-Oncogene-like
46d	CG56698-01	177	178	Mas Proto-Oncogene-like
47	CG56700-01	179	180	Peptidyl-Prolyl Cis-Trans Isomerase-like
48a	CG56743-01	181	182	Phospholipase C Delta-4-like
48b	CG56743-02	183	184	Phospholipase C Delta-4-like
49	CG56739-01	185	186	Leukotriene-B4 Omega-Hydroxylase-like
50a	CG56771-01	187	188	Protein Arginine N-Methyltransferase 2-like
50b	CG56771-02	189	190	Protein Arginine N-Methyltransferase 2-like
51	CG56759-01	191	192	Olfactory Receptor-like
52	CG56731-01	193	194	H326-like
53	CG56745-01	195	196	Uracil Phosphoribosyltransferase-Like
54a	CG56773-01	197	198	Protein Phosphatase 2C-like
54b	CG56773-02	199	200	Protein Phosphatase 2C-like
55	CG56806-01	201	202	Heparan Sulfate 6-Sulfotransferase 3-like
56a	CG56816-01	203	204	N-Hydroxyarylamine Sulfotransferase-like
56b	CG56816-02	205	206	N-Hydroxyarylamine Sulfotransferase-like
57	CG56829-01	207	208	Testis Specific Serine Kinase-3-like
58a	CG56315-01	209	210	Gap Junction Beta-5-like
58b	CG56315-02	211	212	Connexin-like
59	CG56633-01	213	214	Translation Initiation Factor 5-like
60a	CG56894-01	215	216	Lynx1-like
60b	CG56894-02	217	218	Lynx1-like
61	CG56453-01	219	220	Adlcan-like
62	CG56781-01	221	222	Neurotrophin Precursor-like
63	CG53054-02	223	224	Wnt-14 Precursor-like
64	CG56884-01	225	226	Dipeptidyl peptidase-like
65a	CG56651-01	227	228	Protein phosphatase-like
65b	CG56651-02	229	230	Protein phosphatase-like
66	CG56313-01	231	232	Olfactory receptor-like
67	CG56571-01	233	234	Olfactory Receptor-Like Protein OLF3-like
68	CG56844-01	235	236	Endoglin (CD105 antigen)-like

69a	CG56950-01	237	238	Interleukin 1 Epsilon-like
69b	CG56136-02	239	240	Interleukin 1 Epsilon-like
70a	CG56878-01	241	242	OS-9-like
70b	CG56878-04	243	244	OS-9-like
71	CG56906-01	245	246	Sodium/Hydrogen Exchanger 6-like
72	CG56910-01	247	248	Ubiquitin-Specific Protease-like
73	CG56822-01	249	250	Sulfotransferase-like
74	CG56775-01	251	252	Dual Specificity Phosphatase-like
75	CG56783-01	253	254	Dual Specificity Phosphatase-like
76a	CG56789-01	255	256	Dual Specificity Phosphatase-like
76b	CG56789-02	257	258	Dual Specificity Phosphatase-like
77	CG56804-01	259	260	Dual Specificity Phosphatase-like
78	CG56810-01	261	262	Dual Specificity Phosphatase-like
79	CG56862-01	263	264	Dual Specificity Phosphatase-like
80	CG56882-01	265	266	Dual Specificity Phosphatase-like
81a	CG56283-01	267	268	Beta-1,3-Galactosyltransferase-like
81b	CG56283-02	269	270	Beta-1,3-Galactosyltransferase-like
82	CG56983-01	271	272	Peptide YY-like
83	CG56890-01	273	274	G Protein-Coupled Receptor Kinase GRK7-like
84	CG56912-01	275	276	Phospholipase C delta 1-like
85	CG56955-01	277	278	GTPase-Activating Protein-like
86	CG56957-01	279	280	GTPase-Activating Protein-like
87a	CG56886-01	281	282	Rho-GTPase-Activating Protein-like
87b	CG56886-02	283	284	Rho-GTPase-Activating Protein-like
88	CG56394-01	285	286	Glycerol-3-Phosphate Dehydrogenase-like
89	CG56396-01	287	288	Glycerol-3-Phosphate Dehydrogenase-like
90	CG56888-01	289	290	Serine/Threonine-Protein Kinase PAK 2-like
91	CG56779-01	291	292	D-Dopachrome Tautomerase-like
92	CG56904-01	293	294	Secreted leucine-rich repeat (LRR)-like
93	CG56277-01	295	296	Inosine-5'-Monophosphate Dehydrogenase-like
94	CG56281-01	297	298	Male-Specific Lethal 3-Like 1-like
95	CG56975-01	299	300	Cysteine Conjugate Beta-Lyase-like
96a	CG56918-01	301	302	Monocarboxylate transporter-like
96b	CG56918-02	303	304	Monocarboxylate transporter-like
96c	CG56918-03	305	306	Sugar Transporter-like
97a	CG57070-01	307	308	Carboxypeptidase A1-like
97b	CG57070-02	309	310	Carboxypeptidase A1-like
97c	CG57070-03	311	312	Carboxypeptidase A1-like

97d	CG57070-04	313	314	Carboxypeptidase A1-like
97e	CG57070-05	315	316	Carboxypeptidase A1-like
97f	CG57070-06	317	318	Carboxypeptidase A1-like
98	CG56939-01	319	320	Agrin-like
99	CG57010-01	321	322	SNC73-like

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

NOV1, NOV3, and NOV4 are homologous to a Claudin-like family of proteins. Thus, the NOV1, NOV3, and NOV4 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

NOV2 is homologous to the Protein Serine Kinase-like family of proteins. Thus NOV2 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

NOV5 is homologous to a family of Monocarboxylate transporter-like proteins. Thus, the NOV5 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV6 is homologous to the nitrilase-1-like family of proteins. Thus, NOV6 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

NOV7 is homologous to the Cleavage Signal-1-like family of proteins. Thus NOV7 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

NOV8 is homologous to the Matripase-like family of proteins. Thus NOV8 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in, various pathologies or conditions.

NOV9 is homologous to members of the Neuropeptide Y/Peptide YY receptor-like family of proteins. Thus, the NOV9 nucleic acids, polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

5 NOV10 through 20, , NOV43, NOV44, and NOV83 are homologous to the G-Protein Coupled Receptor-like family of proteins. Thus, these nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

10 NOV21 and NOV22 are homologous to the Adrenal; secretory serine protease like growth factor binding protein-like family of proteins. Thus, NOV21 and NOV22 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

15 NOV23 is homologous to the Serine Protease DESC-1-like family of proteins. Thus, NOV23 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in various pathologies or conditions.

NOV24 is homologous to the Pargorin-like family of proteins. Thus, NOV24 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or disorders.

20 NOV25 is homologous to the Protein Phosphatase-like family of proteins. Thus, NOV25 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions.

25 NOV26 is homologous to the GAGE7-like family of proteins. Thus, NOV26 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies/disorders.

30 NOV27 is homologous to the Sodium-Glucose Cotransporter-like family of proteins. Thus, NOV27 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV28 is homologous to the MYD-1-like family of proteins. Thus, NOV28 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV29 is homologous to the CRAL-TRIO-like family of proteins. Thus, NOV27 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

5 NOV30 is homologous to the Ryudocan-like family of proteins. Thus, NOV30 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV31 is homologous to the Sulfur-rich Keratin-like family of proteins. Thus,
10 NOV31 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV32 is homologous to the DMNT1 associated protein-like family of proteins. Thus, NOV32 nucleic acids and polypeptides, antibodies and related compounds according to
15 the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV33 is homologous to the Notch1-like family of proteins. Thus, NOV33 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or
20 conditions

NOV34, NOV35, NOV51, NOV66, and NOV67 are homologous to the Olfactory Receptor-like family of proteins. Thus, NOV34, NOV35, NOV51, NOV66, and NOV67 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or
25 conditions

NOV36 is homologous to the Cadherin 11-like family of proteins. Thus, NOV36 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

30 NOV37 is homologous to the Ten-M2-like family of proteins. Thus, NOV33 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV38 and NOV39 are homologous to the Activin/Inhibin-like family of proteins. Thus, NOV38 and NOV39 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

5 NOV40 is homologous to the UDP Glycosyltransferase-like family of proteins. Thus, NOV40 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

10 NOV41 is homologous to the Sodium/Hydrogen Exchanger 4-like family of proteins. Thus, NOV41 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

15 NOV42 is homologous to the Kupffer Cell Receptor-like family of proteins. Thus, NOV42 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

20 NOV45 and NOV46 is homologous to the Mas Proto-Oncogene-like family of proteins. Thus, NOV45 and NOV46 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV47 is homologous to the Peptidyl-Prolyl Cis-Trans Isomerase-like family of proteins. Thus, NOV47 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

25 NOV48 is homologous to the Phospholipase C Delta-4-like family of proteins. Thus, NOV48 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

30 NOV49 is homologous to the Leukotriene-B4 Omega Hydroxylase-like family of proteins. Thus, NOV49 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV50 is homologous to the Protein Arginine N-Methyltransferase 2-like family of proteins. Thus, NOV50 nucleic acids and polypeptides, antibodies and related compounds

according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV52 is homologous to the H326-like family of proteins. Thus, NOV52 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV53 is homologous to the Uracil Phosphoribosyltransferase-like family of proteins. Thus, NOV53 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV54 is homologous to the Protein Phosphatase 2C-like family of proteins. Thus, NOV54 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV55 is homologous to the Heparan Sulfate 6-Sulfotransferase 3-like family of proteins. Thus, NOV55 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV56 is homologous to the N-Hydroxyarylamine Sulfotransferase 3-like family of proteins. Thus, NOV56 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV57 is homologous to the Testis Specific Serine Kinase-3-like family of proteins. Thus, NOV57 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV58 is homologous to the Gap Junction Beta-5-like family of proteins. Thus, NOV58 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV59 is homologous to the Translation Initiation Factor 5-like family of proteins. Thus, NOV59 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV60 is homologous to the Lynx1-like family of proteins. Thus, NOV60 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

5 NOV61 is homologous to the Adlican-like family of proteins. Thus, NOV61 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

10 NOV62 is homologous to the Neuropsin Precursor-like family of proteins. Thus, NOV62 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

15 NOV63 is homologous to the Wnt-14-like family of proteins. Thus, NOV63 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

20 NOV64 is homologous to the Dipeptidyl peptidase-like family of proteins. Thus, NOV64 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV65 is homologous to the Protein phosphatase-like family of proteins. Thus, NOV65 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

25 NOV68 is homologous to the Endoglin (CD105 antigen)-like family of proteins. Thus, NOV68 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

30 NOV69 is homologous to the Interleukin 1 Epsilon-like family of proteins. Thus, NOV69 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV70 is homologous to the OS-9-like family of proteins. Thus, NOV70 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be

useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV71 is homologous to the Sodium/Hydrogen Exchanger 6-like family of proteins. Thus, NOV71 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV72 is homologous to the Ubiquitin Specific Protease-like family of proteins. Thus, NOV72 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV73 is homologous to the Sulfotransferase-like family of proteins. Thus, NOV73 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV74, NOV75, NOV76, NOV77, NOV78, NOV79, and NOV80 are homologous to the Dual Specificity Phosphatase-like family of proteins. Thus, NOV74, NOV75, NOV76, NOV77, NOV78, NOV79, and NOV80 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV81 is homologous to the Beta-1, 3-Galactosyltransferase-like family of proteins. Thus, NOV81 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV82 is homologous to the Peptide YY-like family of proteins. Thus, NOV82 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV84 is homologous to the Phospholipase C delta 1-like family of proteins. Thus, NOV84 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV85, NOV86, and NOV87 are homologous to the GTPase-Activating Protein-like family of proteins. Thus, NOV85, NOV86, and NOV87 nucleic acids and polypeptides,

antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV88 and NOV89 are homologous to the Glycerol-3-Phosphate Dehydrogenase-like family of proteins. Thus, NOV88 and NOV89 nucleic acids and polypeptides, antibodies
5 and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV90 is homologous to the Serine/Threonine-Protein Kinase PAK 2-like family of proteins. Thus, NOV90 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated
10 in various pathologies or conditions

NOV91 is homologous to the D-Dopachrome Tautomerase family of proteins. Thus, NOV91 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV92 is homologous to the Secreted leucine-rich repeat (LRR)-like family of proteins. Thus, NOV92 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated
15 in various pathologies or conditions

NOV93 is homologous to the Inosine-5'-Monophosphate Dehydrogenase-like family of proteins. Thus, NOV93 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated
20 in various pathologies or conditions

NOV94 is homologous to the Male-Specific Lethal 3-like family of proteins. Thus, NOV94 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various
25 pathologies or conditions

NOV95 is homologous to the Cysteine Conjugate Beta Lyase-like family of proteins. Thus, NOV95 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various
30 pathologies or conditions

NOV96 is homologous to the Monocarboxylate transporter-like family of proteins. Thus, NOV96 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

NOV97 is homologous to the Carboxypeptidase A1-like family of proteins. Thus, NOV97 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

5 NOV98 is homologous to the Agrin-like family of proteins. Thus, NOV98 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

10 NOV99 is homologous to the SNC73-like family of proteins. Thus, NOV99 nucleic acids and polypeptides, antibodies and related compounds according to the invention will be useful in therapeutic and diagnostic applications implicated in various pathologies or conditions

15 The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell proliferation, hematopoiesis, wound healing and angiogenesis.

Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

20 NOV1

NOV1 includes three novel human Claudin-like proteins disclosed below. The disclosed sequences have been named NOV1a, NOV1b, NOV1c, NOV1d, NOV1e, NOV1f, and NOV1g.

25 NOV1a

30 A disclosed NOV1a nucleic acid of 687 nucleotides (also referred to as CG56592-02) encoding a novel human Claudin 6-like protein is shown in Table 1A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 6-8 and ending with a TAG termination codon at nucleotides 678-680. The start and stop codons are in bold letters in Table 1A, and the 5' and 3' untranslated regions are underlined.

Table 1A. NOV1a nucleotide sequence (SEQ ID NO:1).

TGACTATGGCCTGGAGTTTCCGTGCAAAAGTCCAGCTCGGGGGGCTACTTCTCCCTCCTTGGCTGGGTCT
GCTCCTGTGTTACCACCATCCTGCCCCAGTGGAAGACTCTTAATCTGGAAGTGAACGAGATGGAGACCTGGA
TCATGGGGATTGGGAGGTCTGCGTGGATCGAGAGGAAGTCGCCACTGTGTGCAAGGCCTTTGAATCCTTCT
TGTCTCTGCCCCAGGAGCTCCAGGTAGCCCGCATCCTCATGGTAGCCTCCCATGGGCTGGGCCTATTGGGGC
TTTTGCTCTGCAGCTTTGGGTCTGAATGCTTCCAGTTTACAGGATCAGATGGGTATTCAAGAGGCGGCTTG
GTCTCCTGGGAAGGACTTTGGAGGCATCCGCTTCAGCCACTACCCTCCTTCCAGTCTCCTGGGTGGCCCATG
CCACAATCCAAGACTTCTGGGATGACAGCATCCCTGACATCATAACCCTCGGTGGGAGTTTGGAGGTGCCCTC
TACTTGGGCTGGGCTGCTGGTATTTTCTGGCTCTTGGTGGGCTACTCCTCATCTTCTCGGCCTGCCTGGGA
AAAGAAGATGTGCCTTTTCTTTGATGGCTGGTCCACAGTCCCCCTATCCTGTGCTCCAGTGGAGGAGTCA
GATGGCTCCTTCCACCTCATGCTAAGACCTAGGAACCTG

In a search of public sequence databases, the NOV1a nucleic acid sequence, located on chromosome 12 has 337 of 534 bases (63%) identical to a gb:GENBANK-ID:HSA249735|acc:AJ249735.1 mRNA from *Homo sapiens* (CLDN6 gene for claudin-6).

In all BLAST alignments herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the NOV1a BLAST analysis, e.g., *Homo sapiens* CLDN6 gene for claudin-6, matched the Query NOV1a sequence purely by chance is $1.4e^{-15}$. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, e.g.,

<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/>. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than

significant position-by-position alignment. (Wootton and Federhen, Methods Enzymol 266:554-571, 1996).

The disclosed NOV1a polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 has 229 amino acid residues and is presented in Table 1B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1a has no signal peptide and is likely to be localized the plasma membrane with a certainty of 0.6400. Alternatively, NOV1a also may localize to the Golgi body with acertainty of 0.4600, the endoplasmic reticulum (membrane) with a certainty of 0.3700 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV1a peptide is between amino acids 24 and 25, at: VCS-CV.

Table 1B. Encoded NOV1a protein sequence (SEQ ID NO:2).

MAWSFRAKVQLGGLLSLLGWVCSVTTILPQWKTLNLELNEMETWIMGIWEVCVDREEVATVCKAFESFLS LPQELQVARIILMVASHGLGLLGLLLCSFGSECFQFHRIRWVFKRRLGLLGRTLEASASATTLPLVSWVAHAT IQDFWDDSI PDII PRWEFGGALYLGWAAGIFLALGGLLLIFSA CLGKEDVPFPLMAGPTVPLSCAPVEESDG SFHMLRLRPNLVI

A search of sequence databases reveals that the NOV1a amino acid sequence has 81 of 207 amino acid residues (39%) identical to, and 111 of 207 amino acid residues (53%) similar to, the 219 amino acid residue ptnr:SWISSPROT-ACC:Q9Z262 protein from *Mus musculus* (Mouse) (Claudin-6) ($E = 2.7e^{-27}$).

NOV1a is predicted to be expressed in Bone Marrow, Brain, Liver, Placenta, and Lung.

NOV1b

A disclosed NOV1b nucleic acid of 687 nucleotides (also referred to as CG56586-01) encoding a human Claudin-3-like protein is shown in Table 1C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 6-8 and ending with a TAG codon at nucleotides 678-680. Putative untranslated regions upstream from the initiation codon, and downstream from the termination codon, if any, are underlined in Table 1C. The start and stop codons are in bold letters.

Table 1C. NOV1b nucleotide sequence (SEQ ID NO:3).

TGACTATGGCCTGGAGTTTCCGTGCAAAAGTCCAGCTCGGGGGCTACTTCTCTCCCTCCTTGGCTGGGTCT
GCTCCTGTGTTACCACCATCCTGCCCCAGTGGAAGACTCTTAATCTGGAAGTGAACGAGATGGAGACCTGGA
TCATGGGGATTGGGAGGTCTGCGTGGATCGAGAGGAAGTCGCCACTGTGTGCAAGGCCTTTGAATCCTTCT
TGTCTCTGCCCCAGGAGCTCCAGGTAGCCCGCATCCTCATGGTAGCCTCCCATGGGCTGGGCCTATGGGGC
TTTTGCTCTGCAGCTTTGGGTCTGAATGCTTCCAGTTTCACAGGATCAGATGGGTATTCAAGAGGCGGCTTG
GTCTCCTGGGAAGGACTTTGGAGGCATCCGCTTCAGCCACTACCTCCTTCCAGTCTCCTGGGTGGCCCATG
CCACAATCCAAGACTTCTGGGATGACAGCATCCCTGACATCATAACCTCGGTGGGAGTTTGGAGGTGCCCTC
TACTTGGGCTGGGCTGCTGGTATTTTCTGGCTCTTGGTGGGCTACTCCTCATCTTCTCGGCCTGCTGGGA
AAAGAAGATGTGCCTTTTCTTTGATGGCTGGTCCCACAGTCCCCCTATCCTGTGCTCCAGTGGAGGAGTCA
GATGGCTCCTTCCACCTCATGCTAAGACCTAGGAACCTG

In a search of public sequence databases, the NOV1b nucleic acid sequence, located on chromosome 11 is 338 of 534 bases (63%) identical to a gb:GENBANK-

ID:HSA249735|acc:AJ249735.1 mRNA from *Homo sapiens* (CLDN6 gene for claudin-6) (E = 2.8e⁻¹⁶).

The disclosed NOV1b polypeptide (SEQ ID NO:4) encoded by SEQ ID NO:3 has 224 amino acid residues and is presented in Table 1D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1b has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.4600. Alternatively, NOV1b may also localize to the microbody (peroxisome) with a certainty of 0.3200, the endoplasmic reticulum (membrane) with a certainty of 0.1000 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV1b peptide is between amino acids 24 and 25, at: VCS-CV.

Table 1D. Encoded NOV1b protein sequence (SEQ ID NO:4).

MAWSFRAKVQLGGLLSLLGWVCSCVTILPQWKTLNLELNEMETWIMGIWEVCVDREEVATVCKAFESFLS
LPQELQVARILMVASHGLGLLGLLLCSFGSECFQFHRIRWVFKRRLGLLGRTLEASASATLLPVSWVAHAT
IQDFWDDSI PDI I PSVGWVRCP LLGLGCWYFPGSWWATPHLLGLPGKRRCAFSFDGWSHSPILCSSGGVWR
LLPPHAKT

A search of sequence databases reveals that the NOV1b amino acid sequence has 50 of 149 amino acid residues (33%) identical to, and 83 of 149 amino acid residues (55%) similar to, the 219 amino acid residue ptnr:SWISSPROT-ACC:Q63400 protein from *Rattus norvegicus* (Rat) (Claudin-3 (Ventral Prostate.1 Protein) (RVP1)) (E = 0.0).

NOV1b is predicted to be expressed in Bone Marrow, Brain, Liver, Placenta, and Lung.

NOV1c

A disclosed NOV1c nucleic acid of 642 nucleotides (also referred to as CG56592-03) encoding a novel Claudin-6-like protein is shown in Table 1E. An open reading frame was

identified beginning with a ATG initiation codon at nucleotides 6-8 and ending with a TAG codon at nucleotides 609-611. The start and stop codons are in bold letters, and the 5' and 3' untranslated regions are underlined.

Table 1E. NOV1c Nucleotide Sequence (SEQ ID NO:5)

TGACTATGGCTGGAGTTCCGTGCAAAAGTCCAGCTCGGGGGGCTACTTCTCTCCCTCCTTGGCTGGGTC
TGCTCCTGTGTTACCACCATCTGCCCCAGTGAAGACTCTTAATCTGGAAGTGAACGAGATGGAGACCTG
GATCATGGGGATTGGGAGGTCTGCGTGGATCGAGAGGAAGTCGCCACTGTGTGCAAGGCCCTTTGAATCCT
TCTGTCTCTGCCCCAGGAGCTCCAGTTTCACAGGATCAGATGGGTATTCAAGAGGCGGCTTGGTCTCCTG
GGAAGGACTTTGGAGGCATCCGCTTCAGCCACTACCCTCCTTCCAGTCTCCTGGGTGGCCCATGCCACAAT
CCAAGACTTCTGGGATGACAGCATCCCTGACATCATACCTCGGTGGGAGTTTGGAGGTGCCCTCTACTTGG
GCTGGGCTGCTGGTATTTTCCTGGCTCTTGGTGGGCTACTCCTCATCTTCTCGGCCTGCCCTGGGAAAAGAA
GATGTGCCTTTTCCTTTGATGGCTGGTCCACAGTCCCCCTATCCTGTGCTCCAGTGGAGGAGTCAGATGG
CTCCTTCCACCTCATGCTAAGACCTAGGAACCTGGTCATCTAGGACTGGCTTCTGCCAAGGATCTCTGGAA
TAA

The disclosed NOV1c nucleic acid sequence maps to chromosome 12 and has 144 of 220 bases (65%) identical to a gb:GENBANK-ID:HSA249735|acc:AJ249735.1 mRNA from *Homo sapiens* (CLDN6 gene for claudin-6) (E = 0.0).

A disclosed NOV1c protein (SEQ ID NO:6) encoded by SEQ ID NO:5 has 201 amino acid residues, and is presented using the one-letter code in Table 1F. Signal P, Psort and/or Hydropathy results predict that NOV1c does have a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.4600. In other embodiments NOV1c is also likely to be localized to the microbody (peroxisome) with a certainty of 0.2651, to endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV1c is between positions 24 and 25, (VCS-CV).

Table 1F. Encoded NOV1c protein sequence (SEQ ID NO:6).

MAWSFRAKVQLGGLLLSLGWVCSCVTTILPQWKTLLNLELNEMETWIMGIWEVCVDREEVATVCKAFESFL
SLPQELQFHRIRWVFKRRLGLLGRLEASASATLLPVSVAHATIQDFWDDSIPIIIPRWEFGGALYLGW
AAGIFLALGGLLLIFSACLGKEDVPFPLMAGPTVPLSCAPVEESDGSFHLMLRPRNLVI

The disclosed NOV1c amino acid has 55 of 94 amino acid residues (58%) identical to, and 62 of 94 amino acid residues (65%) similar to, the 220 amino acid residue ptnr:SPTREMBL-ACC:Q9D7U6 protein from *Mus musculus* (Mouse) (2210404A22RIK Protein) (E= 3.1e⁻⁴⁷).

In addition, NOV1c is predicted to be expressed in Bone Marrow, Brain, Liver, Placenta, and Lung.

NOV1d

A disclosed NOV1d nucleic acid of 726 nucleotides (also referred to as CG56592-02) encoding a novel Claudin 6-like protein is shown in Table 1G. An open reading frame was identified beginning with an ATG codon at nucleotides 6-8 and ending with a TAG codon at nucleotides 693-695. The start and stop codons are in bold letters and the 5' and 3' untranslated regions are underlined in Table 1G.

Table 1G. NOV1d nucleotide sequence (SEQ ID NO:7).

TGACTATGGCCTGGAGTTTCCGTGCAAAAGTCCAGCTCGGGGGGGCTACTTCTCTCCCTCCTTGGCTGGGTCT
 GTTCCTGTGTTACCAACCATCCTGCCCCAGTGGGAAGACTCTTAATCTGGAAGTGAACGAGATGGAGACCTGGA
 TCATGGGGATTTGGGAGGTCTGCGTGGATCGAGAGGAAGTCGCCACTGTGTGCAAGGCCTTTGAATCCTTCT
 TGTCTCTGCCCCAGGAGCTCCAGGTAGCCCGCATCCTCATGGTAGCCTCCCATGGGCTGGGCCTATTGGGGC
 TTTTGCTCTGCAGCTTTGGGTCTGAATGCTTCCAGTTTCACAGGATCAGATGGGTATTCAAGAGGCGGCTTG
 GTCTCCTGGGAAGGACTTTGGAGGCATCCGCTTCAGCCACTACCTCCTTCCAGTCTCCTGGGTGGCCCATG
 CCACAATCCAAGACTTCTGGGATGACAGCATCCCTGACATCATACCTCGGTGGGAGTTTGGAGGTGCCCTCT
 ACTTGGGCTGGGCTGCTGGTATTTCTGGCTCTTGGTGGGCTACTCCTCATCTTCTCGGCCTGCCTGGGAA
 AAGAAGATGTGCCTTTTCCTTTGATGGCTGGTCCCACAGTCCCCCTATCCTGTGCTCCAGTGGAGGAGTCAG
 ATGGCTCCTTCCACCTCATGCTAAGACCTAGGAACCTGGTCATCTAGGACTGGCTTCTGCCAAGGATCTCTG
GAATAA

In a search of public sequence databases, the NOV1d nucleic acid sequence, located on chromosome 12 has 336 of 534 bases (62%) identical to a gb:GENBANK-

ID:HSA249735|acc:AJ249735.1 mRNA from *Homo sapiens* (CLDN6 gene for claudin-6) (E = 6.5e⁻¹⁶).

The disclosed NOV1d polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 has 229 amino acid residues and is presented in Table 1H using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV1d has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6400. Alternatively, NOV1d may also localize to the Golgi body with a certainty of 0.4600, the endoplasmic reticulum (membrane) with a certainty of 0.3700 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV1d peptide is between amino acids 24 and 25, at: VCS-CV.

Table 1H. Encoded NOV1d protein sequence (SEQ ID NO:8).

MAWSFRAKVQLGGLLSLLGWVCSCVTTLIPQWKTLLNLELNEMETWIMGIWEVCVDREEVATVCKAFESFLS
 LPQELQVARIIMVASHGLGLLGLLCSFGSECFQFHRIRWVFKRRLGLLGRILEASATLLPVSWVAHAT
 IQDFWDDSIPTDIPRWEFGGALYLGAAGIFLALGGLLIIFSACLKEDVPFPLMAGPTVPLSCAPVEESDG
 SFHMLRPRNLVI

A search of sequence databases reveals that the NOV1d amino acid sequence has 81 of 207 amino acid residues (39%) identical to, and 111 of 207 amino acid residues (53%) similar

to, the 219 amino acid residue ptnr:SWISSPROT-ACC:Q9Z262 protein from *Mus musculus* (Mouse) (Claudin-6) ($E = 2.8e^{-27}$).

Expression information was derived from the tissue sources of the sequences that were included in the derivation of NOV1d. The sequence is predicted to be expressed in Bone Marrow, Brain, Liver, Placenta, and Lung.

Homologies to any of the above NOV1 proteins will be shared by the other NOV1 proteins insofar as they are homologous to each other as shown below. Any reference to NOV1 is assumed to refer to all four of the NOV1 proteins in general, unless otherwise noted.

The disclosed NOV1a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 1I.

Table 1I. BLAST results for NOV1a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17458947 ref XP_061964.1 (XM_061964)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	229	229/229 (100%)	229/229 (100%)	e-109
>gi 17437506 ref XP_068031.1 (XM_068031)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	220	99/172 (57%)	125/172 (72%)	4e-50
gi 17437504 ref XP_068030.1 (XM_068030)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	220	99/172 (57%)	126/172 (72%)	4e-43
gi 12843248 dbj BAB25914.1 (AK008821)	PMP-22/EMP/MP20/Claudin family containing protein-data source: Pfam, source key: PF00822, evidence: ISS-putative [<i>Mus musculus</i>]	220	104/188 (55%)	131/188 (69%)	3e-40
gi 7710002 ref NP_057883.1 (NM_016674)	claudin 1 [<i>Mus musculus</i>]	211	67/194 (34%)	99/194 (50%)	2e-20

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 1J. In the ClustalW alignment of the NOV1 proteins, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and

.....

5 1) Novel NOV1a (SEQ ID NO:2)
2) Novel NOV1b (SEQ ID NO:4)
3) Novel NOV1c (SEQ ID NO:6)
4) Novel NOV1d (SEQ ID NO:8)
10 5) gi|17458947|ref|XP_061964.1| (XM_061964) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:1383)
6) gi|17437506|ref|XP_068031.1| (XM_068031) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:1384)
7) gi|17437504|ref|XP_068030.1| (XM_068030) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:325)
15 8) gi|12843248|dbj|BAB25914.1| (AK008821) PMP-22/EMP/MP20/Claudin family containing protein-data source:Pfam, source key:PF00822, evidence:ISS-putative [*Mus musculus*] (SEQ ID NO:326)
9) gi|7710002|ref|NP_057883.1| (NM_016674) claudin 1 [*Mus musculus*] (SEQ ID NO:327)

24

gi 17437504	178	LGGCLLNCAACSSHAPLALGHYAQAQMQ-----TQCPYLEDTADPOV	220
gi 12843248	178	LGGCVLHCAACWSPAPAASSHYAVAGPR-----DHQOHLKQANPET	220
gi 7710002	177	LGGVLLSCSPRKTTSYTPRPYPKPTF-----SSGKDYV	211

5

The claudins are a family of integral membrane proteins that are major components of tight junction (TJ) strands. When claudins are introduced into cells that lack tight junctions, networks of strands and grooves form at cell-cell contact sites that closely resemble native tight junctions. There are at least 17 members of this family in mammals. Claudin family members share ~38% amino acid identity, and are predicted to have four transmembrane (TM) domains, which is reminiscent of occludin, although they share no sequence similarity with it. Multiple sequence alignment reveals their sequences to be fairly well conserved in the first and fourth putative TM domains, and in the first and second extracellular loops, but they diverge in the second and third TM domains. Although the sequences of their C-terminal cytoplasmic domains vary, the known family members share a common motif of -Y-V. This has been postulated as a possible binding motif for PDZ domains of other tight junction-associated peripheral membrane proteins, such as ZO-1.

The disclosed NOV1 nucleic acid of the invention encoding a Human Claudin -like protein includes the nucleic acid whose sequence is provided in Table 1A ,1C, 1E, 1G, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 1A, 1C, 1E, or 1G while still encoding a protein that maintains its Human Claudin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV1 protein of the invention includes the Human Claudin-like protein whose sequence is provided in Table 1B, 1D, 1F, or 1H. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 1B, 1D, 1F, or 1H while still encoding a protein that maintains its Human

Claudin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 66 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

5 The above disclosed information suggests that this Human Claudin-like protein (NOV1) is a member of a "Human Claudin family". Therefore, the NOV1 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small
10 molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV1 nucleic acids and proteins of the invention are useful in potential
15 therapeutic applications implicated in cancer including but not limited to various pathologies and disorders as indicated below. For example, a cDNA encoding the Human Claudin-like protein (NOV1) may be useful in gene therapy, and the Human Claudin-like protein (NOV1) may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients
20 suffering from Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies,
25 Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic lupus erythematosus, Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, and Cancer, or other pathologies or conditions. The NOV1 nucleic acid encoding the Human Claudin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
30 assessed.

 NOV1 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies"

section below. The disclosed NOV1 proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV2

A disclosed NOV2 nucleic acid of 1361 nucleotides (also referred to as CG56596-01) encoding a novel Protein Serine Kinase-like protein is shown in Table 2A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 20-22 and ending with a TAA codon at nucleotides 1268-1270. A The start and stop codons are in bold letters in Table 2A.

Table 2A. NOV2 nucleotide sequence (SEQ ID NO:9).

```
CGGCGGGCGTGTGCGGGTATGGGGTGCGGGCGCCAGCAGGAAGGTGGTCCCGGGGCCACCAAAATTCTTGT
AATAGAATTGGCATCCAAAGTGGAACCCAGAAATGGAACAAAGAATGATCTCTATAAAATTTTTTATTATAC
TTTAAGTTCTACTCCTCCCTGCCCTCTGCCACTCCCCTCACTACCCAGTGCCCCCTCCCTCCTTGCCCTGG
GCCCGAGGCGGCGGGCCAGGCGGCGCAGAGGATACAGGTGGCTCGCTCCGAGCCAAGTTCGACCCCCGGGT
CCTTGCCAGATATGACATCAAAGCTCTTATTGGGACAGGCAGTTTCAGCAGGGTGTGTCAGGGTAGAGCAGAA
GACCACCAAGAAACCTTTTGCAATAAAAGTGATGGAACCCAGAGAGAGGGAAGGTAGAGAAGCGTGCGTGTC
TGAGCTGAGCGTCTTGCGGCGGGTTAGCCATCGTTACATTGTCCAGTTCATGGAGATCTTTGAGACTGAGGA
TCAAGTTTACATGGTAATGGAGCTGGCTACCGAGGGGAGCTCTTTGATCGACTCATTGCTCAGGGATCCTT
TACAGAGCGGGATGCCGTCAGGATCCTCCAGATGGTTGCTGATGGGATTAGGTATTTGCATGCGCTGCAGAT
AACTCATAGGAATCTAAAGCCTGAAAACCTCTTATACTATCATCCAGGTGAAGAGTCGAAAATTTTAATTAC
AGATTTTGGTTTGGCATACTCCGGGAAAAAAGTGGTGACTGGACAATGAAGACACTCTGTGGGACCCCCAGA
GTACATAGCTCCTGAGGTTTGGCTAAGGAAGCCTTATACAGTGCACTGGACATGTGGGCTCTTGGTGTGAT
CACATATGCTTTACTTAGCGGATTCCTGCCTTTTGATGATGAAAGCCAGACAAGGCTTTACAGGAAGATTCT
GAAAGGCCAAATATAATTATACAGGAGAGCCTTGGCCAAGCATTCCCACTTGCGGAAGGACTTTATAGACAA
ACTACTGATTTTGGAGGCTGGTCATCGCATGTCAGCTGGCCAGGCCCTGGACCATCCCTGGGTGATCACCAT
GGCTGCAGGGTCTTCCATGAAGAATCTCCAGAGGGCCATATCCCGAAACCTCATGCAGAGGGCCTCTCCCA
CTCTCAGAGTCTGGATCTGCACAGTCTTCTAAGTCACATTATTCTCACAAATCCAGGCATATGTGGAGCAA
GAGAACTTAAGGATAGTAGAATCGCCACTGTCTGCGCTTTTGTTAAGCAGATGACCTCTAAACTATTTTTG
CCTATTTTAGGACCATTTCATCATGATTAGGGCACCTCAAGCTCCAAAGACACGGGACTCCATG
```

The disclosed NOV2 nucleic acid sequence, localized to the q21.3-22 region of chromosome 18, has 685 of 997 bases (68%) identical to a gb:GENBANK-ID:HSA272212|acc:AJ272212.1 mRNA from *Homo sapiens* (mRNA for protein serine kinase (PSKH1 gene)) ($E = 6.1e^{-85}$).

A NOV2 polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 has 416 amino acid residues and is presented using the one-letter code in Table 2B. Signal P, Psort and/or Hydropathy results predict that NOV2 contains no signal peptide and is likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.5500. Alternatively, NOV2 may also localize to the lysosome (lumen) with a certainty of 0.2403, the plasma membrane with a certainty of 0.1900, or the microbody (peroxisome) with a certainty of 0.1111.

Table 2B. Encoded NOV2 protein sequence (SEQ ID NO:10).

MGCGASRKVVPKILVIELASKVEPRNGTKNDLYKFFYYTLLSSTPPCPLPLPSLPQCPLPPCPGPEAAAQ
AAQRIQVARFRAKFDPRVLARYDIKALIGTGSFVRVVEQKTTKKPFAIKVMETREEREACVSELSVLR
RVSHRYIVQLMEIFETEDQVYVMELATGGELFDRLIAQGSFTERDAVRILQMVDGIRYLHALQITHRNK
PENLLYYHPGEESKILITDFGLAYSGKKSGDWTMKTLCGTPEYIAPEVLLRKPYTSAVDMWALGVITYALLS
GFLPFDDSESQTRLRKILKGKYNITGEPWPSISHLAKDFIDKLLILEAGHRMSAGQALDHPWVITMAAGSSM
KNLQRAISRNLQMRASPHSQSPGSAQSSKSHYSHKSRHMWSKRNLRIVESPLSALL

The disclosed NOV2 amino acid sequence has 267 of 412 amino acid residues (64%) identical to, and 332 of 412 amino acid residues (80%) similar to, the 424 amino acid residue ptnr:SPTREMBL-ACC:Q9NY19 protein from *Homo sapiens* (Human) (Protein Serine Kinase) ($E = 1.1e^{-138}$).

NOV2 is predicted to be expressed in Kidney, Lymph node, Pancreas, Salivary Glands, Brain, and Placenta because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSA272212|acc:AJ272212.1) a closely related *Homo sapiens* mRNA for protein serine kinase (PSKH1 gene) homolog.

In addition, the sequence is predicted to be expressed in keratinocytes because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSPI13711|acc:AJ001696.2) a closely related *Homo sapiens* mRNA for hurpin, clone R7-1.1 homolog.

NOV2 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 2C.

Table 2C. BLAST results for NOV2

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14916455 ref NP_149117.1 (NM_033126)	serine/threonine kinase PSKH2 [<i>Homo sapiens</i>]	385	369/416 (88%)	372/416 (88%)	0.0
gi 17530179 gb AAL40735.1 (AF416988)	protein serine kinase/luciferase fusion protein	975	257/391 (65%)	318/391 (80%)	e-149
gi 14776113 ref XP_043047.1 (XM_043047)	hypothetical protein XP_043047 [<i>Homo sapiens</i>]	424	257/391 (65%)	318/391 (80%)	e-145
gi 15963448 gb AAL11033.1 (AF236367)	protein serine kinase Pskh1 [<i>Mus musculus</i>]	424	254/386 (65%)	311/386 (79%)	e-144

			
	NOV2	416	-----	416
	gi 14916455	385	-----	385
5	gi 17530179	719	FFAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLPGIRQGYGLTETTSAILITP	778
	gi 14776113	424	-----	424
	gi 15963448	424	-----	424
	gi 2136035	319	-----	319
			790 800 810 820 830 840	
10	NOV2	416	416
	gi 14916455	385	-----	385
	gi 17530179	779	EGDDKPGAVGKVVPFPEAKVVDLDTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALID	838
	gi 14776113	424	-----	424
15	gi 15963448	424	-----	424
	gi 2136035	319	-----	319
			850 860 870 880 890 900	
20	NOV2	416	416
	gi 14916455	385	-----	385
	gi 17530179	839	KDGWLHSGDIAYWDEDEHFFIVDRLSLIKYGQVAPAELESILLQHPNIFDAGVAGLP	898
	gi 14776113	424	-----	424
	gi 15963448	424	-----	424
25	gi 2136035	319	-----	319
			910 920 930 940 950 960	
30	NOV2	416	416
	gi 14916455	385	-----	385
	gi 17530179	899	DDDAGELPAAVVVLEHGKTMTEKEIVDYVASQVTTAKKLRGGVVFDEVPKGLTGKLDAR	958
	gi 14776113	424	-----	424
	gi 15963448	424	-----	424
35	gi 2136035	319	-----	319
			970	
	NOV2	416	416
	gi 14916455	385	-----	385
40	gi 17530179	959	KIREILIKAKKGGKIAV	975
	gi 14776113	424	-----	424
	gi 15963448	424	-----	424
	gi 2136035	319	-----	319
45				

The presence of identifiable domains in NOV2, as well as all other NOVX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>).

DOMAIN results for NOV2 as disclosed in Tables 2E-2G, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 2K and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading or by the sign (!) and “strong” semi-conserved residues are indicated by grey shading or by the sign (+). The “strong” group of conserved amino acid

residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Tables 2E-G lists the domain description from DOMAIN analysis results against NOV2. This indicates that the NOV2 sequence has properties similar to those of other proteins known to contain this domain.

Table 2E Domain Analysis of NOV2

	NOV 2:	94	YDIKALIGTGSFSRVRVEQKTTKKPFAIKVMETRE--REGREACVSELSVLRVRVSHRYI	151
10	Sbjct:	1	YELLEVLGKGAFGKVYLARDKKTGKLVAIKVIKKEKLKKKKRERILREIKILKKLDHPNI	60
	NOV 2:	152	VQLMEIFETEDQVYVMELATGGELFDRLIAQGSFTERDAVRILQMVDGIRYLHALQIT	211
15	Sbjct:	61	VKLYDVFEDEDDKLYLVMMEYCEGGDLFDLLKKRGRLESEARFYARQILSALEYLHSQGII	120
	NOV 2:	212	HRNLKPENLLYYHPGEESKILITDFGLAYSGKKSQDWTMTKTLCGTPEYIAPEVLLRKPYT	271
20	Sbjct:	121	HRDLKPENILLDSGDH--VKLADFGLA-KQLDSGGTLLTTFVGTPEYMAPEVLLGKGYG	176
	NOV 2:	272	SAVDMWALGVITYALLSGFLPFDDDESQTRLYRKILKGYNYTGEPWPSISHLAKDFIDKL	331
25	Sbjct:	177	KAVDIWSLGVILYELLTGKPPFPGDDQLLALFKKIGKPPPPFPPPEWKISPEAKDLIKKL	236
	NOV 2:	332	LILEAGHRMSAGQALDHPWV	351
25	Sbjct:	237	LVKDPEKRLTAEEALEHPFF	256

25

20

25

Table 2F Domain Analysis of NOV2

30	NOV 2:	94	YDIKALIGTGSFSRVVRVEQKTTKKPFAIKVMETREREGREACV-SELSVLRVRVSHRYIV	152
	Sbjct:	1	YELGEKLGSGAFGKVYKGKHKDTGEIVAIIKILKKRSLSEKKKRFLREIQILRRLSHPNIV	60
35	NOV 2:	153	QLMEIFETEDQVYVMELATGGELFDRLIAQGS-FTERDAVRILQMVADGIRYLHALQIT	211
	Sbjct:	61	RLLGVFEEDDHLVLYVMEYMEGGDLFDYLRNRGLLLSEKEAKKIALQILRGLEYLHSGRIV	120
40	NOV 2:	212	HRNLKPENLLYYHPGEESKILITDFGLAYSGKKSQDWTMTKLTGTPEYIAPEVLLRKPYT	271
	Sbjct:	121	HRDLKPENILLDENGTT--VKIADFGFLARKLESSSYEKLTTTFVGTPEYMAPEVLEBGRGYS	177
45	NOV 2:	272	SAVDMWALGVITYALLSGFLPFDDDESQTRLYRKILKGKYNITGEPWPSISHLAKDFIDKL	331
	Sbjct:	178	SKVDVWSLGVILYELLTGKLPPFGIDPLEELFRIKERPR-LRLPLPNCSEELKDIIKKC	236

35

40

45

Table 2G Domain Analysis of NOV2

gnl|Smart|smart00219, TyrKc, Tyrosine kinase, catalytic domain;
Phosphotransferases. Tyrosine-specific kinase subfamily. (SEQ ID
NO:801)
CD-Length = 258 residues, 83.7% aligned
Score = 117 bits (292), Expect = 2e-27

5	NOV 2:	100	IGTGSFSRVVR---VEQKTTKKPFAIKVM-ETREREGREACVSELSVLRRVSHRYIVQLM	155
	Sbjct:	7	LGEGAFGEVYKGTTLKGGVEVEVAVKTLKEDASEQQIEEFLREARLMRKLDHPNIVKLL	66
	NOV 2:	156	EIFETEDQVYMVMELATGGELFDRLIAQG--SFTERDAVRILQMVADGIRYLHALQITHR	213
	Sbjct:	67	GVCTEEELMIVMEYMEGGDLLDYLRKNRPKELSLDLLSFALQIARGMEYLESKNFVHR	126
10	NOV 2:	214	NLKPENLLYYHPGEESKILITDFGLAYSGKKSGDWTMKTLCGTP-EYIAPEVLLRKPYTS	272
	Sbjct:	127	DLAARNCLV---GENKTVKIADFGLARDLYDDDYRKKKSPRLPIRWMAPESLKDGKFTS	183
15	NOV 2:	273	AVDMWALGVITYALLS-GFLPFDESQTRLYRKILKGY	310
	Sbjct:	184	KSDVWSFGVLLWEIFTLGESPYPGMSNEEVLEYLKKGYR	222

Protein phosphorylation is a fundamental process for the regulation of cellular functions. The coordinated action of both protein kinases and phosphatases controls the levels of phosphorylation and, hence, the activity of specific target proteins. One of the predominant roles of protein phosphorylation is in signal transduction, where extracellular signals are amplified and propagated by a cascade of protein phosphorylation and dephosphorylation events. Eukaryotic protein kinases are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common with both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme.

The disclosed NOV2 nucleic acid of the invention encoding a Protein Serine Kinase-like protein includes the nucleic acid whose sequence is provided in Tables 2A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Tables 2A while still encoding a protein that maintains its Protein Serine Kinase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include

chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
5 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 32 percent of the bases may be so changed.

The disclosed NOV2 protein of the invention includes the Protein Serine Kinase -like protein whose sequence is provided in Tables 2B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown
10 in Table 2B while still encoding a protein that maintains its Protein Serine Kinase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 35 percent of the residues may be so changed.

The NOV2 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Diabetes, Von Hippel-Lindau (VHL) syndrome ,
15 Pancreatitis, Obesity, Lymphedema , Allergies, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney
20 disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, and/or other pathologies and disorders.

NOV2 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the
25 art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which are useful in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV3

NOV3 includes three novel human Claudin-like proteins disclosed below. The disclosed sequences have been named NOV3a, NOV3b, and NOV3c.

NOV3a

A disclosed NOV3a nucleic acid of 695 nucleotides (designated CuraGen Acc. No. CG56594-01) encoding a novel Claudin-19-like protein is shown in Table 3A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 53-55 and ending with a TGA codon at nucleotides 662-664. A putative untranslated region downstream from the termination codon is underlined in Table 3A, and the start and stop codons are in bold letters.

Table 3A. NOV3a Nucleotide Sequence (SEQ ID NO:11)

GCACCCCTGGCCAGCTCTGAGTCCTGGGACCCTCGGTCTCTCTCCTGGGCCC**ATGG**CCAACTCAGGCCTC
CAGCTCCTGGGCTACTTCTTGGCCCTGGGTGGCTGGGTGGGCATCATTTGCTAGCACAGCCCTGCCACAGT
GGAAGCAGTCTTCCTACGCAGGCGACGCCATCATCACTGCCGTGGGCCTCTATGAAGGGCTCTGGATGTC
CTGCGCCTCCAGAGCACTGGGCAAGTGCAAGCTCTACGACTCGCTGCTCGCCCTGGACGGTAGG
CCCCAGGCCCGCGGGCCCTGATGGTGGTGGCCGTGCTCCTGGGCTTCGTGGCCATGGTCTCAGCGTAG
TTGGCATGAAGTGACGCGGTGGGAGACAGCAACCCATTGCCAAGGGCCGTGTGCCATCGCCGGGGG
AGCCCTCTTCATCTGGCAGGCCTCTGCACTTTGACTGCTGTCTCGTGGTATGCCACCCTGGTGACCCAG
GAGTTCTTCAACCCAGAATTTGGCCAGCCCTGTTCTGGGCTGGGCCTCAGCTGGCCTGGCCGTGCTGG
GCGGCTCCTTCCTCTGCTGCACATGCCCCGAGCCAGAGAGACCAACAGCAGCCACAGCCCTATCGGCC
TGGACCCTCTGCTGCTGCCCCGAGAGTACGTCT**GAG**CTCCGCTGCCCTGGCCAGCCCCCACCCA

The nucleic acid sequence, localized to chromosome 1, has 402 of 482 bases (83%) identical to a gb:GENBANK-ID:AF249889|acc:AF249889.1 mRNA from *Mus musculus* (claudin-19 mRNA, partial cds) ($E = 1.1e^{-67}$).

A NOV3a polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 is 203 amino acid residues and is presented using the one letter code in Table 3B. Signal P, Psort and/or Hydropathy results predict that NOV3a has no signal peptide and is likely to be localized at the endoplasmic reticulum (membrane) with a certainty of 0.6850. Alternatively, NOV3a may also localize to the plasma membrane with a certainty of 0.6400, the Golgi body with a certainty of 0.4600, or the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV3a is between positions 23 and 24: IIA-ST.

Table 3B. NOV3a protein sequence (SEQ ID NO:12)

MANSGQLLLGYFLALGGWVGIIASTALPQWKQSSYAGDAITAVGLYEGLWMSCASQSTGQVQCKLYDSLALD
GRPQAARALMVAVLLGFVAMVLSVVGMMKCTRVGDSNPVIAKGRVAIAGGALFILAGLCTLTAVSWYATLVTQEF
FNPEFGPALFVGWASAGLAVLGGSFLCCTCPEPERPNSSPQPYRPGPSAAAREYV

The full amino acid sequence of the protein of the invention was found to have 174 of 193 amino acid residues (90%) identical to, and 178 of 193 amino acid residues (92%) similar to, the 193 amino acid residue ptnr:TREMBLNEW-ACC:AAF98323 protein from *Mus musculus* (Mouse) (CLAUDIN-19) ($E = 5.7e^{-89}$).

NOV3a is predicted to be expressed in at least the Spinal cord.

NOV3b

A disclosed NOV3b nucleic acid of 695 nucleotides (also referred to as CG56594-01) encoding a novel Claudin-19-like protein is shown in Table 3C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 53-55 and ending with a TGA termination codon at nucleotides 662-664. The start and stop codons are in bold letters in Table 3C, and the 5' and 3' untranslated regions are underlined.

Table 3C. NOV3b nucleotide sequence (SEQ ID NO:13).

```
GCACCCTGGCCAGCTCTGAGTCCTGGGACCCTCGGTCTCTCTCTGGGCCATGGCCAACCTCAGGCCTCCA
GCTCTCTGGGCTACTTCTTGGCCCTGGGTGGCTGGGTGGGCATCATTGCTAGCACAGCCCTGCCACAGTGGAA
GCAGTCTTCCCTACGCAGGCGACGCCATCATCACTGCCGTGGGCCTCTATGAAGGGCTCTGGATGTCCTGCCG
CTCCAGAGCACTGGGCAAGTGCAAGCTCTACGACTCGCTGCTCGCCCTGGACGGTAGGCCCCAGGC
CGCGCGGGCCCTGATGGTGGTGGCCGTGCTCTCTGGGCTTCGTGGCCATGGTCCTCAGCGTAGTTGGCATGAA
GTGTACGCGGGTGGGAGACAGCAACCCCATTGCCAAGGGCGGTGTTGCCATCGCCGGGGAGCCCTCTTCAT
CCTGGCAGGCCTCTGCACTTTGACTGCTGTCTCGTGGTATGCCACCCTGGTGACCCAGGAGTTCTTCAACCC
AGAATTTGGCCAGCCCTGTTCTGTGGCTGGGCCTCAGCTGGCCTGGCCGTGCTGGGCGGCTCCTCTCTGTG
CTGCACATGCCCGGAGCCAGAGAGACCCAAACAGCAGCCACAGCCCTATCGGCCTGGACCCTCTGCTGTCTGC
CCGAGAGTACGTCTGAGCTCCGCCTGCCCTGGCCAGCCCCCACCCA
```

In a search of public sequence databases, the NOV3b nucleic acid sequence, located on chromosome 1 has 402 of 482 bases (83%) identical to a gb:GENBANK-ID:AF249889|acc:AF249889.1 mRNA from *Mus musculus* (claudin-19 mRNA, partial cds) (E = $1.1e^{-67}$).

The disclosed NOV3b polypeptide (SEQ ID NO:14) encoded by SEQ ID NO:13 has 203 amino acid residues and is presented in Table 3D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV3b has a signal peptide and is likely to be localized the endoplasmic reticulum (membrane) with a certainty of 0.6850. Alternatively, NOV3b may also localize to the plasma membrane with a certainty of 0.6400, the Golgi body with a certainty of 0.4600 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV3b is between positions 23 and 24: IIA-ST.

Table 3D. Encoded NOV3b protein sequence (SEQ ID NO:14).

```
MANSGLQLLYFLALGGWVGI IASTALPQWKQSSYAGDAIITAVGLYEGLWMSCASQSTGQVQCKLYDSLILA
LDGRPQAARALMVVAVLLGFVAMVLSVGMKCTRVGDSNPIAKGRVAIAGGALFILAGLCTLTAVSWYATLV
TQEFFNPEFGPALFVGWASAGLAVLGGSFLLCTCPEPERPNSSPQPYRPGPSAAAREYV
```

A search of sequence databases reveals that the NOV3b amino acid sequence has 174 of 193 amino acid residues (90%) identical to, and 178 of 193 amino acid residues (92%)

similar to, the 193 amino acid residue ptnr:TREMBLNEW-ACC:AAF98323 protein from *Mus musculus* (Mouse) (Claudin-19) ($E = 5.7e^{-89}$).

NOV3b is predicted to be expressed in at least the Spinal cord.

NOV3c

5 A disclosed NOV3c nucleic acid of 690 nucleotides (also referred to as CG57576-01) encoding a novel Claudin 19-like protein is shown in Table 3E. An open reading frame was identified beginning with an ATG codon at nucleotides 51-53 and ending with a TGA codon at nucleotides 684-686. The start and stop codons are in bold letters and the 5' and 3' untranslated regions are underlined in Table 3I. Because the start codon is not a traditional
10 initiation codon, NOV3c could be a partial reading frame. NOV3c could extend further in the 5' direction.

Table 3E. NOV3c nucleotide sequence (SEQ ID NO:15).

<p> ACCCTGGCCCAGTCTCTGAGTCCTGGGACCCTCGGTCCTCTCTCCTGGGCCATGGCCAACTCAGGCCTCCAGC TCCTGGGCTACTTCTTGGCCCTGGGTGGCTGGGTGGGCATCATTGCTAGCACAGCCCTGCCACAGTGGAAGC AGTCTTCCTACGCAGGCGACGCCATCATCACTGCCGTGGGCTCTATGAAGGGCTCTGGATGTCTTGGCCCT CCCAGAGCACTGGGCAAGTGCAAGCTCTACGACTCGCTGCTCGCCCTGGACGGTCACATCCAATCAG CGCGGGCCCTGATGGTGGTGGCCGTGCTCCTGGGCTTCGTGGCCATGGTCCTCAGCGTAGTTGGCATGAAGT GTACGCGGGTGGGAGACAGCAACCCATTGCCAAGGGCCGTGTTGCCATCGCCGGGGAGCCCTCTTCATCC TGGCAGGCCCTCTGCACTTTGACTGCTGTCTCGTGGTATGCCACCCTGGTGACCCAGGAGTTCTTCAACCCAA GCACACCTGTCAATGCCAGGTATGAATTTGGCCAGCCCTGTTTCGTGGGCTGGGCCTCAGCTGGCCTGGCCG TGCTGGGCGGCTCCTTCCTCTGCTGCACATGCCCGGAGCCAGAGACCCAAACAGCAGCCACAGCCCTATC GGCCTGGACCCTCTGCTGCTGCCCGAGAGTACGCTTGAGCTC </p>
--

15 In a search of public sequence databases, the NOV3c nucleic acid sequence, located on chromosome 1 has 445 of 671 bases (66%) identical to a gb:GENBANK-ID:HSA011497|acc:AJ011497.1 mRNA from *Homo sapiens* (mRNA for Claudin-7) ($E = 5.3e^{-46}$).

20 The disclosed NOV3c polypeptide (SEQ ID NO:16) encoded by SEQ ID NO:15 has 211 amino acid residues and is presented in Table 3F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV3c has no signal peptide and is likely to be localized the the endoplasmic reticulum (membrane) with a certainty of 0.6850. Alternatively, NOV3c may also localize to the plasma membrane with acertainty of 0.6400, the Golgi body with a certainty of 0.4600 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV3c peptide is between amino acids
25 23 and 24, at: IIA-ST.

Table 3F. Encoded NOV3c protein sequence (SEQ ID NO:16).

MANSGQLLLGYFLALGGWVGIIASTALPQWKQSSYAGDAIITAVGLYEGLWMSCASQSTGQVQCKLYDSL
LDGHIQSARALMVAVLLGFVAMVLSVVGKCTRVGDSNPIAKGRVAIAGGALFILAGLCTLTAVSWYATLV
TQEFNPSTPVNARYEFGPALFVGWASAGLAVLGGSFLLCTCPEPERPNSSPQPYRPGPSAAAREYV

A search of sequence databases reveals that the NOV3c amino acid sequence has 121 of 211 amino acid residues (57%) identical to, and 159 of 211 amino acid residues (75%) similar to, the 211 amino acid residue ptr:SWISSNEW-ACC:O95832 protein from *Homo sapiens* (Human) (Claudin-1 (Senescence-Associated Epithelial Membrane Protein)) ($E = 9.6e^{-66}$).

NOV3c is predicted to be expressed in at least Spinal cord.

NOV3a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 3G.

Table 3G. BLAST results for NOV3a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 9789476 gb AAF98323.1 (AF249889)	claudin-19 [<i>Mus musculus</i>]	193	174/193 (90%)	178/193 (92%)	1e-84
gi 17489134 ref XP_060892.1 (XM_060892)	similar to claudin-19 (H. sapiens) [<i>Homo sapiens</i>]	309	126/137 (91%)	127/137 (91%)	3e-59
gi 12654455 gb AAH01055.1 AAH01055 (BC001055)	claudin 7 [<i>Homo sapiens</i>]	211	112/211 (53%)	149/211 (70%)	2e-55
gi 10835008 ref NP_001298.1 (NM_001307)	claudin 7; Clostridium perfringens enterotoxin receptor-like 2; claudin 9 [<i>Homo sapiens</i>]	211	111/211 (52%)	148/211 (69%)	7e-55
gi 7710002 ref NP_057883.1 (NM_016674)	claudin 1 [<i>Mus musculus</i>]	211	112/212 (52%)	149/212 (69%)	8e-55

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 3H.

Table 3H ClustalW Analysis of NOV3

- 1) NOV3a (SEQ ID NO:12)
- 2) NOV3b (SEQ ID NO:14)
- 3) NOV3c (SEQ ID NO:16)
- 4) gi|9789476|gb|AAF98323.1| (AF249889) claudin-19 [*Mus musculus*] (SEQ ID NO:333)
- 5) gi|17489134|ref|XP_060892.1| (XM_060892) similar to claudin-19 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:334)
- 6) gi|12654455|gb|AAH01055.1|AAH01055 (BC001055) claudin 7 [*Homo sapiens*] (SEQ ID NO:335)
- 7) gi|10835008|ref|NP_001298.1| (NM_001307) claudin 7; Clostridium perfringens enterotoxin receptor-like 2; claudin 9 [*Homo sapiens*] (SEQ ID NO:336)

8) gi|7710002|ref|NP_057883.1| (NM_016674) claudin 1 [*Mus musculus*] (SEQ ID NO:327)

			10	20	30	40	50	60	
5	NOV3A	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
	NOV3b	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
	NOV3c	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
	gi 9789476	1	-----	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	50
	gi 17489134	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
10	gi 12654455	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
	gi 10835008	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
	gi 7710002	1	MANSGLQLLG	YFLALGGWVGII	I	ASTALPQWKQSSYAGDAI	I	TAVGLYEGLWMSCASQSTG	60
			70	80	90	100	110	120	
15	NOV3A	61	QVQCKLYDSLLALD	-----	-----	-----	-----	-----	74
	NOV3b	61	QVQCKLYDSLLALD	-----	-----	-----	-----	-----	74
	NOV3c	61	QVQCKLYDSLLALD	-----	-----	-----	-----	-----	74
	gi 9789476	51	QVQCKLYDSLLALD	-----	-----	-----	-----	-----	64
20	gi 17489134	61	QVQCKLYDSLLALD	ALPPTKPEGTLELPSAAWEDGPS	PRGSPQARKQQQ	SESYRKS	LOG	-----	120
	gi 12654455	61	MMSCKMYDSVLALS	-----	-----	-----	-----	-----	74
	gi 10835008	61	MMSCKMYDSVLALS	-----	-----	-----	-----	-----	74
	gi 7710002	61	QVQCKLYDSLLALD	-----	-----	-----	-----	-----	74
25			130	140	150	160	170	180	
	NOV3A	74	-----	-----	-----	-----	GRPQAARA	-----	82
	NOV3b	74	-----	-----	-----	-----	GRPQAARA	-----	82
	NOV3c	74	-----	-----	-----	-----	CHTQSARA	-----	82
30	gi 9789476	64	-----	-----	-----	-----	CHTQSARA	-----	72
	gi 17489134	121	RVENGVDSEGD	SHPTLS	SSCPGCHSSL	SCTQKAPTQPPSPPARLSVIPPTL	CHTQSARA	-----	180
	gi 12654455	74	-----	-----	-----	-----	AALQATRA	-----	82
	gi 10835008	74	-----	-----	-----	-----	AALQATRA	-----	82
	gi 7710002	74	-----	-----	-----	-----	STLQATRA	-----	82
35			190	200	210	220	230	240	
	NOV3A	83	LMVVAVLLG	FVAMVLSVVG	MKCTR	RVGDSNPTAK	GRVATAGGALFILAGL	CTLTAVSWYAT	142
	NOV3b	83	LMVVAVLLG	FVAMVLSVVG	MKCTR	RVGDSNPTAK	GRVATAGGALFILAGL	CTLTAVSWYAT	142
	NOV3c	83	LMVVAVLLG	FVAMVLSVVG	MKCTR	RVGDSNPTAK	GRVATAGGALFILAGL	CTLTAVSWYAT	142
	gi 9789476	73	LMVVAVLLG	FVAMVLSVVG	MKCTR	RVGDSNPTAK	SRVATAGGALFILAGL	CTLTAVSWYAT	132
	gi 17489134	181	LMVVAVLLG	FVAMVLSVVG	MKCTR	RVGDSNPTAK	GRVATAGGALFILAGL	CTLTAVSWYAT	240
	gi 12654455	83	LMVVSLVLG	FVAMVLSVVG	MKCTR	CGGDDKVKK	KARIAMGGGII	FTVAGLATLVACSWYGH	142
	gi 10835008	83	LMVVSLVLG	FVAMVLSVVG	MKCTR	CGGDDKVKK	KARIAMGGGII	FTVAGLATLVACSWYGH	142
45	gi 7710002	83	LMVIGILLGL	IADFVSTIG	MKCMRCL	EDDEVQKMMMAVIGGII	FLISGLATLVATAWYGN	-----	142
			250	260	270	280	290	300	
50	NOV3A	143	LVTQEFFNP	-----	EF	GPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 192
	NOV3b	143	LVTQEFFNP	-----	EF	GPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 192
	NOV3c	143	LVTQEFFNP	STPNARYE	FGPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 200	
	gi 9789476	133	LVTQEFFNP	STPNARYE	FGPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 190	
	gi 17489134	241	LVTQEFFNP	STPNARYE	FGPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 298	
	gi 12654455	143	QIVTD	FYNPLIPTN	IKYEF	GPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 202
55	gi 10835008	143	QIVTD	FYNPLIPTN	IKYEF	GPALFVGWASAGLAVL	GGSF	LCCTCP	EPERPNSSPQPYR 202
	gi 7710002	143	RIVQEFYD	ELTPIN	ARYEFG	QALFTGWAASL	CLLGGV	LLSCSCE	RKTTSYPTPRPYR 200
			310						
60	NOV3A	193	PGPSAAAREYV	-----	-----	-----	-----	-----	203
	NOV3b	193	PGPSAAAREYV	-----	-----	-----	-----	-----	203
	NOV3c	201	PGPSAAAREYV	-----	-----	-----	-----	-----	211
	gi 9789476	191	SGP	-----	-----	-----	-----	-----	193
	gi 17489134	299	PGPSAAAREYV	-----	-----	-----	-----	-----	309
65	gi 12654455	203	K--SNSSKEYV	-----	-----	-----	-----	-----	211
	gi 10835008	203	K--SNSSKEYV	-----	-----	-----	-----	-----	211
	gi 7710002	201	KPTPSSGKDYV	-----	-----	-----	-----	-----	211

just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
5 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 17 percent of the bases may be so changed.

The disclosed NOV3 protein of the invention includes the Claudin-19 -like protein
10 whose sequence is provided in Table 3B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 3B while still encoding a protein that maintains its Claudin-19 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 48 percent of the residues may be so changed.

The protein similarity information, expression pattern, and map location for the
15 Claudin-19-like protein and nucleic acid (NOV3) disclosed herein suggest that this NOV3 protein may have important structural and/or physiological functions characteristic of the Claudin-19 family. Therefore, the NOV3 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific
20 or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition
25 promoting tissue regeneration in vitro and in vivo.

The NOV3 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Hemophilia,
30 hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft versus host, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic lupus

erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, and Cancer, and/or other pathologies. The NOV3 nucleic acids, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

5 NOV3 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. These novel proteins can be used in assay systems for functional analysis of
10 various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV4

NOV4 includes three novel human I Claudin-like proteins disclosed below. The
15 disclosed sequences have been named NOV4a, NOV4b, and NOV4c.

NOV4a

A disclosed NOV4a nucleic acid of 694 nucleotides (also referred to as CG56589-01) encoding a novel Claudin-6-like protein is shown in Table 4A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 11-13 and ending with a
20 TAA codon at nucleotides 671-673. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 4A, and the start and stop codons are in bold letters.

Table 4A. NOV4a Nucleotide Sequence (SEQ ID NO:17)

ACCTGTCGCAATGGCTTTAATCTTTAGAACAGCAATGCAATCTGTTGGACTTTTACTATCTC
TCCTGGGATGGATTTTATCCATTATTACAACCTATTTGCCACACTGGAAGAACCTCAACCTG
GACTTAAATGAAATGGAAACTGGACCATGGGACTCTGGCAAACCTGTGTCTATCCAAGAGGA
AGTGGGGATGCAATGCAAGGACTTTGACTCCTTCCTGGCTTTGCCTGCTGAACTCAGGGTCT
CCAGGATCTTAATGTTTCTGTCAAATGGGCTGGGATTTCTGGGCCTGCTGGTCTCTGGGTTT
GGCCTGGACTGTTTGAGAATTGGAGAGAGTCAGAGAGATCTCAAGAGGCGACTGCTCATTCT
GGGAGGAATTCGTCTGGGCCTCGGGAATCACAGCCCTGGTTCCCGTCTCTTGGGTTGCC
ACAAGACGGTTCAGGAGTTCCTGGGATGAGAACGTCCCAGACTTTGTCCCCAGGTGGGAGTTT
GGGGAGGCCCTGTTTCTGGGCTGGTTTGCTGGACTTTCTCTCTGCTAGGAGGGTGTCTGCT
CAACTGCGCAGCCTGCTCCAGCCACGCTCCCCTAGCTTTGGGCCACTATGCAGTGGCGCAAA
TGCAAACCTCAGTGTCCCTACCTGGAAGATGGGACAGCAGATCCTCAAGTG**TAA**GA**CTCCGAC**
AAGGCCAGAGAT

The NOV4a nucleic acid was identified on chromosome 4 and has 330 of 556 bases (59%) identical to a gb:GENBANK-ID:AF134160|acc:AF134160.1 mRNA from *Homo sapiens* (claudin-1 (CLDN1) mRNA, complete cds) ($E = 2.9e^{-9}$).

A disclosed NOV4a polypeptide (SEQ ID NO:18) encoded by SEQ ID NO:17 is 220 amino acid residues and is presented using the one-letter code in Table 4B. Signal P, Psort and/or Hydropathy results predict that NOV4a has no signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6400. Alternatively, NOV4a may also localize to the Golgi body with a certainty of 0.4600, the endoplasmic reticulum (membrane) with a certainty of 0.3700, or the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV4a is between positions 24 and 25: ILS-II.

Table 4B. Encoded NOV4a protein sequence (SEQ ID NO:18)

MALIFRTAMQSVGLLLSLLGWILSIITTYLPHWKNLNLDLNEMENWTMGLWQTCVIQEEVGMQCKDFDSFLA LPAELRVSRILMFLSNGLGFLGLLVSGFGLDCLRIGESQRDLKRRLILGGILSWASGITALVPVSWVAHKT VQEFWDENVPDFVPRWEFGEALFLGWFAGLSLLGGCLLNCAACSSHAPLALGHYAVAQMTCPCPYLEDGTA DPQV
--

The disclosed NOV4a amino acid sequence has 84 of 204 amino acid residues (41%) identical to, and 119 of 204 amino acid residues (58%) similar to, the 219 amino acid residue ptnr:SWISSPROT-ACC:Q9Z262 protein from *Mus musculus* (Mouse) (Claudin-6) ($E = 1.1e^{-32}$).

NOV4a is predicted to be expressed in at least Brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in Adrenal Gland/Suprarenal gland, Brain, Bronchus, Brown adipose, Cervix, Colon, Coronary Artery, Epidermis, Gall Bladder, Heart, Hippocampus, Islets of Langerhans, Kidney, Liver, Lung, Lung Pleura, Mammary gland/Breast, Oesophagus, Ovary, Oviduct/Uterine Tube/Fallopian tube, Parotid Salivary glands, Peripheral Blood, Placenta, Prostate, Proximal Convolutd Tubule, Respiratory Bronchiole, Skin, Stomach, Substantia Nigra, Thymus, Thyroid, Trachea, Umbilical Vein, Uterus, and Vulva.

NOV4b

A disclosed NOV4b nucleic acid of 694 nucleotides (also referred to as CG56589-01) encoding a novel Claudin-6-like protein is shown in Table 4C. An open reading frame was identified beginning with an ATG codon at nucleotides 11-13 and ending with a TAA codon at

nucleotides 671-673. The start and stop codons are in bold letters and the 5' and 3' untranslated regions are underlined in Table 4C. Because the start codon is not a traditional initiation codon, NOV4b could be a partial reading frame. NOV4b could extend further in the 5' direction.

Table 4C. NOV4b nucleotide sequence (SEQ ID NO:19).

ACCTGTCGCAATGGCTTTAATCTTTAGAACAGCAATGCAATCTGTTGGACTTTTACTATCTCTCTGGGATG
GATTTTATCCATTATTACAACCTATTGCGCACACTGGAAGAACCTCAACCTGGACTTAAATGAAATGAAAA
CTGGACCATGGGACTCTGGCAAACCTGTGTCATCCAAGAGGAAGTGGGGATGCAATGCAAGGACTTTGACTC
CTTCTCGGCTTTGCCTGCTGAACTCAGGGTCTCCAGGATCTTAATGTTTCTGTCAAATGGGCTGGGATTTCT
GGCCTGCTGGTCTCTGGGTTTGGCCTGGACTGTTTGAGAATTGGAGAGAGTCAGAGAGATCTCAAGAGGCG
ACTGCTCATTCTGGGAGGAATTCTGTCTGGGCTCGGGAATCACAGCCCTGGTTCCCGTCTCTGGGTTGC
CCACAAGACGGTTCAGGAGTCTGGGATGAGAACGTCCCAGACTTTGTCCCCAGGTGGGAGTTTGGGGAGGC
CCTGTTTCTGGGCTGGTTTGCTGGACTTTCTCTTCTGCTAGGAGGGTGTCTGCTCAACTGCGCAGCCTGCTC
CAGCCACGCTCCCCTAGCTTTGGGCCACTATGCAGTGGCGCAAATGCAAACCTCAGTGTCCCTACCTGGAAGA
TGGGACAGCAGATCCTCAAGT**TAAGACTCCGACAAGGCCAGAGAT**

In a search of public sequence databases, the NOV4b nucleic acid sequence, located on chromosome 4 has 330 of 556 bases (59%) identical to a gb:GENBANK-ID:AF134160|acc:AF134160.1 mRNA from *Homo sapiens* (claudin-1 (CLDN1) mRNA, complete cds) ($E = 2.9e^{-09}$).

The disclosed NOV4b polypeptide (SEQ ID NO:20) encoded by SEQ ID NO:19 has 220 amino acid residues and is presented in Table 4D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV4b has no signal peptide and is likely to be localized the the plasma membrane with a certainty of 0.6400. Alternatively, NOV4b may also localize to the Golgi body with acertainty of 0.4600, the endoplasmic reticulum (membrane) with a certainty of 0.3700 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV4b peptide is between amino acids 24 and 25, at: ILS-II.

Table 4D. Encoded NOV4b protein sequence (SEQ ID NO:20).

MALIFRTAMQSVGLLLSLLGWILSIITTYLPHWKNLNDLNEMENWTMGLWQTCVIOEEVGMQCKDFDSFLA
LPALRLVSRILMFLSNGLGFLGLLVSGFGLDCLRIGESQRDLKRRLILGGILSWASGITALVPVSWVAHKT
VQEFWDENVPDFVPRWEFGALFLGWFAGLSLLGGCLLNCAACSSHAPLALGHYAVAQMOTQCPYLEDGTA
DPQV

A search of sequence databases reveals that the NOV4b amino acid sequence has 84 of 204 amino acid residues (41%) identical to, and 119 of 204 amino acid residues (58%) similar to, the 219 amino acid residue ptnr:SWISSPROT-ACC:Q9Z262 protein from *Mus musculus* (Mouse) (Claudin-6) ($E = 1.1e^{-32}$).

NOV4b is predicted to be expressed in at least Brain.

In addition, NOV4b is predicted to be expressed in Adrenal Gland/Suprarenal gland, Brain, Bronchus, Brown adipose, Cervix, Colon, Coronary Artery, Epidermis, Gall Bladder, Heart, Hippocampus, Islets of Langerhans, Kidney, Liver, Lung, Lung Pleura, Mammary gland/Breast, Oesophagus, Ovary, Oviduct/Uterine Tube/Fallopian tube, Parotid Salivary glands, Peripheral Blood, Placenta, Prostate, Proximal Convoluted Tubule, Respiratory Bronchiole, Skin, Stomach, Substantia Nigra, Thymus, Thyroid, Trachea, Umbilical Vein, Uterus, and Vulva.

NOV4c

A disclosed NOV4c nucleic acid of 694 nucleotides (also referred to as CG56589-02) encoding a novel Claudin 6-like protein is shown in Table 4E. An open reading frame was identified beginning with an ATG codon at nucleotides 11-13 and ending with a TAA codon at nucleotides 671-673. The start and stop codons are in bold letters and the 5' and 3' untranslated regions are underlined in Table 4E.

Table 4E. NOV4c nucleotide sequence (SEQ ID NO:21).

ACCTGTCGCAATGGCTTTAATCTTTAAAAACAGCAATGCAATCTGTTGGACTTTTGCTATCTTTCCTGGGATG
GATTTTATCCATTATTACAACCTTATTGCCACACTGGAAGAACCTCAACCTGGACTTAAATGAAATGGAAAA
CTGGACCATGGGACTCTGGCAAACCTGTGTTCATCCAAGAGGAAGTGGGGATGCAATGCAAGGACTTTGACTC
CTTCCTGGCTTTGCTGCTGAACCTCAGGGTCTCCAGGATCTTAATGTTTCTGTCAAATGGGCTGGGATTTCT
GGGCTGCTGGTCTCTGGGTTTGGCCTGGACTGTTTGAGAATTGGAGAGAGTCAGAGAGATCTCAAGAGGCG
ACTGCTCATTCTGGGAGGAATTCTGTCTGGGCCTCGGGAATCACGGCCCTGGTTCCCGTCTCTTGGGTTGC
CCACAAGACGGTTTCAGGAGTTCTGGGATGAGAACGTCCAGACTTTGTCCCCAGGTGGGAGTTTGGGGAGGC
CCTGTTTCTGGGCTGGCTTGCTGGACTTTCTCTTCTGCTAGGAGGGTGTCTGCTCAACTGCGCAGCCTGCTC
CAGCCACGCTCCCTAGCTTTGGGCCACTATGCAGTGGCGCAAATGCAAACCTCACTGTCCCTACCTGGAAGA
TGGGACAGCAGATCCTCAAGTGTAAGACTCCGACAAGGCCAGAGAT

In a search of public sequence databases, the NOV4c nucleic acid sequence, located on chromosome 4 has 331 of 556 bases (59%) identical to a gb:GENBANK-ID:AF134160|acc:AF134160.1 mRNA from *Homo sapiens* (claudin-1 (CLDN1) mRNA, complete cds) ($E = 3.2e^{-9}$).

The disclosed NOV4c polypeptide (SEQ ID NO:22) encoded by SEQ ID NO:21 has 220 amino acid residues and is presented in Table 4F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV4c has no signal peptide and is likely to be localized the the plasma membrane with a certainty of 0.6400. Alternatively, NOV4c may also localize to the Golgi body with acertainty of 0.4600, the endoplasmic reticulum (membrane) with a certainty of 0.3700 or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for a NOV4c peptide is between amino acids 24 and 25, at: ILS-II.

Table 4F. Encoded NOV4c protein sequence (SEQ ID NO:22).

MALIFKTAMQSVGLLLSFLGWILSIITTYLPHWKNLNLNEMENWTMGLWQTCVIEEVGMQCKDFDSFLA
LPAELRVSRILMFLSNGLGFLGLLVSGFGLDCLRIGESQRDLKRRLLILGGILSWASGITALVPVSWVAHKT
VQEFWDENVPDFVPRWEFGALFLGWLAGLSLLGGCLLNCAACSSHAPLALGHYAVAQMOTHCPLYLEDGTA
DPQV

A search of sequence databases reveals that the NOV4c amino acid sequence has 83 of 204 amino acid residues (40%) identical to, and 118 of 204 amino acid residues (57%) similar to, the 219 amino acid residue ptrn:SWISSPROT-ACC:Q9Z262 protein from *Mus musculus* (Mouse) (Claudin-6) ($E = 9.6e^{-66}$).

The sequence is predicted to be expressed in the following tissues : Adrenal Gland/Suprarenal gland, Brain, Bronchus, Brown adipose, Cervix, Colon, Coronary Artery, Epidermis, Gall Bladder, Heart, Hippocampus, Islets of Langerhans, Kidney, Liver, Lung, Lung Pleura, Mammary gland/Breast, Oesophagus, Ovary, Oviduct/Uterine Tube/Fallopian tube, Parotid Salivary glands, Peripheral Blood, Placenta, Prostate, Proximal Convoluted Tubule, Respiratory Bronchiole, Skin, Stomach, Substantia Nigra, Thymus, Thyroid, Trachea, Umbilical Vein, Uterus, and Vulva.

NOV4 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 4G.

Table 4G. BLAST results for NOV4

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17437504 ref XP_068030.1 (XM_068030)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	220	220/220 (100%)	220/220 (100%)	e-105
gi 17437506 ref XP_068031.1 (XM_068031)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	220	192/212 (90%)	198/212 (92%)	9e-96
gi 12843248 dbj BAB25914.1 (AK008821)	PMP-22/EMP/MP20/Claudin family containing protein-data source: Pfam, source key: PF00822, evidence: ISS-putative [<i>Mus musculus</i>]	220	158/220 (71%)	182/220 (81%)	3e-70
gi 17458947 ref XP_061964.1 (XM_061964)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	229	108/188 (57%)	137/188 (72%)	2e-45
gi 7710002 ref NP_057883.1 (NM_016674)	claudin 1 [<i>Mus musculus</i>]	211	72/181 (39%)	105/181 (57%)	1e-27

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 4H.

Table 4H Clustal W Sequence Alignment

5

1) NOV4a (SEQ ID NO:18)

2) NOV4b (SEQ ID NO:20)

3) NOV4c (SEQ ID NO:22)

4) gi|17437504|ref|XP_068030.1| (XM_068030) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:325)

10

5) gi|17437506|ref|XP_068031.1| (XM_068031) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:324)

6) gi|12843248|dbj|BAB25914.1| (AK008821) PMP-22/EMP/MP20/Claudin family containing protein~data source:Pfam, source key:PF00822, evidence:ISS-putative [*Mus musculus*] (SEQ ID NO:326)

15

7) gi|17458947|ref|XP_061964.1| (XM_061964) similar to putative (H. sapiens) [*Homo sapiens*] (SEQ ID NO:323)

8) gi|7710002|ref|NP_057883.1| (NM_016674) claudin 1 [*Mus musculus*] (SEQ ID NO:327)

20

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

1

1

1

1

1

1

1

.....

.....

.....

.....

.....

.....

10

20

30

40

50

60

1

1

1

1

1

1

1

.....

.....

.....

.....

.....

.....

117

117

117

117

117

117

117

57

57

57

57

57

57

57

30

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

58

58

58

58

58

58

58

.....

.....

.....

.....

.....

.....

.....

70

80

90

100

110

120

58

58

58

58

58

58

58

.....

.....

.....

.....

.....

.....

.....

117

117

117

117

117

117

117

40

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

118

118

118

118

118

118

118

.....

.....

.....

.....

.....

.....

.....

130

140

150

160

170

180

118

118

118

118

118

118

118

.....

.....

.....

.....

.....

.....

.....

177

177

177

177

177

177

177

50

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

190

200

210

220

230

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

220

220

220

220

220

220

211

55

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

190

200

210

220

230

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

220

220

220

220

220

220

211

60

NOV4a

NOV4b

NOV4c

gi|17437504|

gi|17437506|

gi|12843248|

gi|17458947|

gi|7710002|

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

190

200

210

220

230

178

178

178

178

178

178

177

.....

.....

.....

.....

.....

.....

.....

220

220

220

220

220

220

211

Table 4I lists the domain description from DOMAIN analysis results against NOV4. This indicates that the NOV4 sequence has properties similar to those of other proteins known to contain this domain.

Table 4I Domain Analysis of NOV4

gnl|Pfam|pfam00822, PMP22_Claudin, PMP-22/EMP/MP20/Claudin family.
(SEQ ID NO:802)
CD-Length = 162 residues, 67.3% aligned
Score = 35.0 bits (79), Expect = 0.004

5	NOV 4:	49	GLWQTCVIQEEVGM-QCKDFDSFLALPAELRVSRIIMFLSNGLGFLGLLVSGFGLDCLRI	107
			+ + + + +	
	Sbjct:	41	GLWRNCTTQSGTGQISCKVL---ELNDALQAVQALMILSIILGIISLIVFFFQLFTMRK	96
10	NOV 4:	108	GESQRDLKRRLILGGILSWASGITALVPVSWVAHKTVQEFWDENVPDFVPRWEFGALF	167
			+ + + + ++ +	
	Sbjct:	97	GGR-----FKLAGIIFLVSGLCVLVGASIYTSRIATDF--GNPFTPNRKYSFGYSFI	146
15	NOV 4:	168	LGW	170
	Sbjct:	147	LGW	149

The claudins are a family of integral membrane proteins that are major components of tight junction (TJ) strands. When claudins are introduced into cells that lack tight junctions, networks of strands and grooves form at cell-cell contact sites that closely resemble native tight junctions. There are at least 17 members of this family in mammals. Claudin family members share ~38% amino acid identity, and are predicted to have four transmembrane (TM) domains, which is reminiscent of occludin, although they share no sequence similarity with it. Multiple sequence alignment reveals their sequences to be fairly well conserved in the first and fourth putative TM domains, and in the first and second extracellular loops, but they diverge in the second and third TM domains. Although the sequences of their C-terminal cytoplasmic domains vary, the known family members share a common motif of -Y-V. This has been postulated as a possible binding motif for PDZ domains of other tight junction-associated peripheral membrane proteins, such as ZO-1.

The disclosed NOV4 nucleic acid of the invention encoding a Claudin-6 -like protein includes the nucleic acid whose sequence is provided in Table 4A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 4A while still encoding a protein that maintains its Claudin-6 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or

complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 41 percent of the bases may be so changed.

The disclosed NOV4 protein of the invention includes the Claudin-6 -like protein whose sequence is provided in Table 4B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 4B while still encoding a protein that maintains its Claudin-6-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 61 percent of the residues may be so changed.

The NOV4 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, Cirrhosis, Transplantation, Hemophilia, hypercoagulation, Idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, Graft vesus host, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Systemic lupus erythematosus , Autoimmune disease, Asthma, Emphysema, Scleroderma, allergy, and Cancer, and/or other pathologies and disorders of the like. The NOV4 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV4 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV4 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV5

NOV5 includes three novel Monocarboxylate transporter (MCT3)-like proteins disclosed below. The disclosed sequences have been named NOV5a, NOV5b, NOV5c, NOV5d, and NOV5e.

NOV5a

A disclosed NOV5a nucleic acid of 1502 nucleotides (also referred to as CG56635-01) encoding a novel Monocarboxylate transporter (MCT3)-like protein is shown in Table 5a. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 24-26 and ending with a TGA codon at nucleotides 1365-1367. The start and stop codons are in bold letters in Table 5A.

Table 5A. NOV5a Nucleotide Sequence (SEQ ID NO:23)

```

GTTTCCCAACCCCGCAGACGGCGATGACCCCCAGCCCGCCGACCCCGGATGGGGGCTGGGGCTGGGT
GGTGGCGGCCCGCAGCCTTCGCGATAAACGGGCTGTCTACGGGCTGTGCGCTCGCTGGGCTTGCCTTC
CCTGACCTTGCCGAGCACTTTGACCGAAGCGCCAGGACACTGCGTGGATCAGCGCCCTGGCCTTGGCCG
TGCAGCAGGCAGCCAGCCCGTGGGCGAGCGCCCTGAGCACGCGCTGGGGGGCCCGCCCGTGGTGATGGT
TGGGGGCGTCTCGCCTCGCTGGGCTTCGTCTTCTCGGCTTCGCCAGCGATCTGCTGCATCTCTACCTC
GGCCTGGGCTCCTCGCTGGCTTTGGTTGGGCGCTGGTGTTCGCCCCCGCCCTAGGCACCCCTCTCGCGTT
ACTTCTCCCGCGCTCGAGTCTTGGCGGTGGGGCTGGCGCTCACCGGCAACGGGGCTCCTCGCTGCTCCT
GGCGCCCGCCTTGCAGCTTCTTCTCGATACTTTCGGCTGGCGGGGCGCTCTGCTCCTCCTCGGCGCGATC
ACCCCTCCACCTACCCCCCTGTGGCGCCCTGCTGCTACCCCTGGTTCCTTCTGGAGACCCCCAGCCCCAC
CGCTAGTCCCTAGCTGCCCTCGGCCAGAGTCTGTTCACACGCCGGGCTTCTCAATCTTTGCTCTAGG
CACAGCCCTGGTTGGGGGCGGGTACTTCGTTTCTTACGTGCACTTGGCTCCCCACGCTTTAGACCGGGGC
CTGGGGGGATACGGAGCAGCGCTGGTGGTGGCGTGGCTGCGATGGGGGATGCGGGCGCCCGCTGGTCT
GCGGGTGGCTGGCAGACCAAGGCTGGGTGCCCTCCCCGCGCTCCTGGCCGATTCGGGGCTCTGACTGG
GCTGGGGCTGTGGGTGGTGGGGCTGGTGCCCGTGGTGGCGCGCAAGAGAGCTGGGGGGTCCCTGCTG
GCCGCGGCTGTGGCCTATGGGCTGAGCGCGGGGAGTTACGCCCGCTGGTTCGTTGCTACTCCCGGGC
TGGTGGGCGTCGGAGGTGTGGTGACGCCACAGGGCTGGTGATGATGCTGATGAGCCTCGGGGGCTCCT
GGGCCCTCCCTGTGAGGCTTCTTAAGGGATGAGACAGGAGACTTACCGCCTCTTTCCTCCTGTCTGGT
TCTTTGATCCTCTCCGGCAGCTTCACTACATAGGGTTGCCAGGGCGCTGCCCTCCTGTGGTCCAGCCT
CCCTCCAGCCACGCTCCCCAGAGACGGGGAGCTGCTCCCGCTCCCGAGGCAGTCTTGTGTCCCC
AGGAGGCCCTGGCTCCACTCTGGACACCACTTGTGATTATTTCTTGTGTTGAGCCCTCCCAATAAA
GAATTTTATCGGGTTTCTGAAACCTCCAACGTTCACCAATCTAGGACCCTGAAAATATTCTACATA
AGACAGCCAGAAAGGCTGGTTCAAAGGAACAG

```

The disclosed NOV5a nucleic acid sequence, located on chromosome 17, has 672 of 1110 bases (60%) identical to a gb:GENBANK-ID:AF132610|acc:AF132610.1 mRNA from

Homo sapiens (monocarboxylate transporter MCT3 mRNA, complete cds) ($E = 1.6e^{-29}$).

A disclosed NOV5a polypeptide (SEQ ID NO:24) encoded by SEQ ID NO:23 is 447 amino acid residues and is presented using the one-letter amino acid code in Table 5B. Signal P, Psort and/or Hydropathy results predict that NOV5a contains no signal peptide and is likely to be localized in the endoplasmic reticulum (membrane) with a certainty of 0.6850.

Alternatively, NOV5a is also likely to be localized to the plasma membrane with a certainty of

0.6400, to the Golgi body with a certainty of 0.4600, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000

Table 5B. Encoded NOV5a protein sequence (SEQ ID NO:24).

MTPQPAGPPDGGWGWVAAAAFAINGLSYGLLRSLGLAFPDLAEHFDRSAQDTAWISALALAVQQAASPVGSALS TRWGARPVVMVGGVLAASLGFVFSASFASDLHLHLGLGLLAGFGWALVFAPALGTLTRYFSRRRVLAAGLALTGNG ASSLLLAPALQLLLDTFGWRGALLLLGAITLHLTPCGALLPLVLPDPPAPPRSPLAALGQSLFTRRAFSIFAL GTALVGGGYFVPYVHLAPHALDRGLGGYGAALVVAAMGDAGARLVCGWLADQGWVPLPRLLAVFGALTGLGLW VVGGLVPVVGGEESWGGPLAAAVAYGLSAGSYAPLVFGVLPGLVGVGGVVQATGLVMMMLSLGGLLGPPLSGFLR DETGDFTASFLLSGSLILSGSFIYIGLPRALPSCGPASPPATPPPETGELLPAQAVLLSPGGPGSTLDTTC
--

The disclosed NOV5a amino acid sequence has 96 of 198 amino acid residues (48%) identical to, and 122 of 198 amino acid residues (61%) similar to, the 504 amino acid residue ptnr:SPTREMBL-ACC:O95907 protein from *Homo sapiens* (Human) (DJ1039K5.2 (Similar To Monocarboxylate Transporter (MCT3))) ($E = 1.2e^{-67}$).

NOV5a is predicted to be expressed in at least Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain - whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, retina, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus.

NOV5b

A disclosed NOV5b nucleic acid of 611 nucleotides (also referred to as CG56635-02) encoding a novel Monocarboxylate transporter 3-like protein is shown in Table 5C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 6-8 and ending with a TGA codon at nucleotides 500-502. The start and stop codons are in bold letters in Table 5B.

Table 5C. NOV5b Nucleotide Sequence (SEQ ID NO:25)

ACGGCGATGACCCCCAGCCCGCCGACCCCGGATGGGGGCTGGGGCTGGGTGGTGGCGGCCGAGCCT TCGCGATAAACGGGCTGTCTACGGGCTGCTGCGCTCGCTGGGCCTTGCTTCCCTGACCTTGCCGAGCA CTTTGACCGAAGCGCCAGGACACTGCGTGGATCAGCGCCCTGGCCCTGGCCGTGCAGCAGGCAGCCAGC CCCGTGGGCAGCGCCTGAGCACGCGCTGGGGGGCCCGCCCGTGGTGTGGTGGGGGCGTCTCGCCT CGCTGGGCTTCGTCTTCTCGGCTTTCGCCAGCGATCTGCTGCATCTCTACCTCGGCCTGGGCCCTCCTCGC TGGCTTCTTAAGGATGAGACAGGAGACTTCACCGCCTCTTCTCTCTGTTCTTTGATCCTCTCC GGCAGCTTCATCTACATAGGGTTGCCAGGGCGCTGCCCTCCTGTGGTCCAGCCTCCCTCCAGCCAGC CTCCCCAGAGACGGGGAGCTGCTTCCCGCTCCCGAGGAGTCTTGCTGTCCCCAGAGGCCCTGGCTC CACTCTGGACACCACTTGTGATTATTTCTTGTGAGCCCTCCCCAC
--

The disclosed NOV5b nucleic acid sequence, located on chromosome 17, has 323 of 520 bases (62%) identical to a gb:GENBANK-ID:AF132610|acc:AF132610.1 mRNA from *Homo sapiens* (monocarboxylate transporter MCT3 mRNA, complete cds) ($E = 3.2e^{-18}$).

A disclosed NOV5b polypeptide (SEQ ID NO:26) encoded by SEQ ID NO:25 is 191 amino acid residues and is presented using the one-letter amino acid code in Table 5D. Signal P, Psort and/or Hydropathy results predict that NOV5b contains no signal peptide and is likely to be localized in the endoplasmic reticulum (membrane) with a certainty of 0.9325.

5 Alternatively, NOV5b is also likely to be localized to the plasma membrane with a certainty of 0.4960, to the microbody (peroxisome) with a certainty of 0.3200, or to the Golgi body with a certainty of 0.1900 The most likely cleavage site for NOV5b is between positions 38 and 39: GLA-FP.

Table 5D. Encoded NOV5b protein sequence (SEQ ID NO:26).

MTPQPAGPPDGGWGVVAAAAFAINGLSYGLLRSLGLAFPDLAEHFDRSAQDTAWISALALAVQQAASPVGSALS TRWGARPVVMVGVLASLGFVFSAFASDLLHLHLGLLAGFLRDETGDFTASFLLSGSLILSGSFIYIGLPRAL PSCGPASPPATPPPETGELLPAQAVLLSPGGPGSTLDTTC
--

10 The disclosed NOV5b amino acid sequence has 53 of 110 amino acid residues (48%) identical to, and 72 of 110 amino acid residues (65%) similar to, the 504 amino acid residue ptnr:SPTREMBL-ACC:Q9UBE2 protein from *Homo sapiens* (Human) (Monocarboxylate Transporter MCT3) ($E = 2.9e^{-28}$).

15 NOV5b is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

NOV5c

20 A disclosed NOV5c nucleic acid of 704 nucleotides (also referred to as CG56635-03) encoding a novel Monocarboxylate transporter 3-like protein is shown in Table 5E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 28-30 and ending with a TGA codon at nucleotides 673-675. The start and stop codons are in bold letters in Table 5E.

Table 5E. NOV5c Nucleotide Sequence (SEQ ID NO:27)

CGAGCAGCCAGAGGCTGGATCTCAGGGATGCCAGCTCCCCAGCGGAAGCACAGGCGTGGAGGCTTCTCTC ACAGATGTTTCCCCACCCCGCAGACGGCGATGACCCCCAGCCCGCCGACCCCGGATGGGGGCTGGGG CTGGGTGGTGGCGGCCGAGCCTTCGCGATAAACGGGCTGTCTACGGGCTGCTGCGCTCGCTGGGCCTT GCCTTCCCTGACCTTGCCGAGCACTTTGACCGAAGCGCCAGGACACTGCGTGGATCAGCGCCCTGGCCC TGGCCGTGCAGCAGGCAGCCAGCCCGTGGGCAGCGCCCTGAGCACGCGCTGGGGGGCCCCCGCCGTGGT GATGGTTGGGGGCGTCTCGCCTCGCTGGGCTTCGTCTTCTCGGCTTTCGCCAGCGATCTGCTGCATCTC
--

TACCTCGGCCTGGGCCTCCTCGCTGGCTTCCTAAGGGATGAGACAGGAGACTTCACCGCCTCTTTCCTCC
 TGTCTGGTTCTTTGATCCTCTCCGGCAGCTTCATCTACATAGGGTTGCCAGGGCGCTGCCCTCCTGTGG
 TCCAGCCTCCCTCCAGCCACGCCTCCCCAGAGACGGGGGAGCTGCTTCCCGCTCCCCAGGCAGTCTTG
 CTGTCCCCAGGAGGCCCTGGCTCCACTCTGGACACCACTTGTGATTATTTTCTTGTTTGAGCCCCTCCC
 CCAC

The disclosed NOV5c nucleic acid sequence, located on chromosome 17, has 340 of 547 bases (62%) identical to a gb:GENBANK-ID:AF019111|acc:AF019111.2 mRNA from *Mus musculus* (monocarboxylate transporter 3 (MCT3) mRNA, complete cds) ($E = 2.4e^{-15}$).

A disclosed NOV5c polypeptide (SEQ ID NO:28) encoded by SEQ ID NO:27 is 215 amino acid residues and is presented using the one-letter amino acid code in Table 5F. Signal P, Psort and/or Hydropathy results predict that NOV5c contains no signal peptide and is likely to be localized in the endoplasmic reticulum (membrane) with a certainty of 0.8500.

Alternatively, NOV5c is also likely to be localized to the microbody (peroxisome) with a certainty of 0.6400, to the plasma membrane with a certainty of 0.4400, or to the nucleus with a certainty of 0.3000

Table 5F. Encoded NOV5c protein sequence (SEQ ID NO:28).

MPAPQRKHRRGGFSHRCFPTPQTAMTPQAPGPPDGGWGWVAAAAFAINGLSYGLLRSLGLAFPDIAEHFDRSAQ
 DTAWISALALAVQQAASPVGSALSTRWGARPVVMVGGVLASLGFVFSAFASDLLHLYLGLGLLAGFLRDETGDFT
 ASFLLSGSLILSGSFYIYGLPRALPSCGPASPPATPPPETGELLPAQAVLLSPGGPGSTLDTTTC

The disclosed NOV5c amino acid sequence has 53 of 110 amino acid residues (48%) identical to, and 72 of 110 amino acid residues (65%) similar to, the 504 amino acid residue ptnr:SPTREMBL-ACC:Q9UBE2 protein from *Homo sapiens* (Human) (Monocarboxylate Transporter MCT3) ($E = 2.9e^{-28}$).

NOV5c is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

NOV5d

A disclosed NOV5d nucleic acid of 1513 nucleotides (also referred to as CG56635-04) encoding a novel Monocarboxylate transporter 3-like protein is shown in Table 5G. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 28-30 and ending with a TGA codon at nucleotides 1444-1446. The start and stop codons are in bold letters in Table 5G.

Table 5G. NOV5d Nucleotide Sequence (SEQ ID NO:29)

CGAGCAGCCAGAGGCTGGATCTCAGGGATGCCAGCTCCCCAGCGGAAGCACAGGCGTGGAGGCTTCTCTC
ACAGATGTTTTCCCCACCCCGCAGACGGCGATGACCCCCAGCCCGCCGACCCCGGATGGGGGCTGGGG
CTGGGTGGTGGCGGCCGAGCCTTCGCGATAAACGGGCTGTCTACGGGCTGTGCGCTCGCTGGGCCTT
GCCTTCCCTGACCTTGCCGAGCACTTTGACCGAAGCGCCAGGACACTGCGTGGATCAGCGCCCTGGCCC
TGGCCGTGCAGCAGGCAGCCAGTCCCGTGGGACGCGCCCTGAGCACGCGCTGGGGGGCCCGCCCGTGGT
GATGGTTGGGGGCGTCTCGCCTCGCTGGGCTTCGTCTTCTCGGCTTCGCCAGCGATCTGCTGCATCTC
TACCTCGGCCTGGGCCTCCTCGCTGGTTTTGGTTGGGCCCTGGTGTTCGCCCCGCGCTAGGCACCCCTC
CGCGTTACTTCTCCCGCCGTCGAGTCTTGGCGGTGGGGCTGGCGCTACCGGCAACGGGGCCCTCCTCGCT
GCTCTGGCGCCCGCCTTGCAGCTTCTTCTCGATACTTTCGGCTGGCGGGGCGCTCTGCTCCTCCTCGGC
GCGATCACCTCCACCTACCCCTGTGGCGCCCTGTGCTACCCCTGGTCTTCTTGGAGACCCCGCAG
CCCCACCGCGTAGTCCCTAGCTGCCCTCGGCCTGAGTCTGTTCACACGCGGGGCTTCTCAATCTTTCG
TCTAGGCACAGCCCTGGTTGGGGGGCGGTACTTCGTTCCCTTACGTGCACTTGGCTCCCGACGCTTAGAC
CGGGGCTGGGGGATACGGAGCAGCGCTGGTGGTGGCCGTGGCTGCGATGGGGGATGCGGGCGCCCGGC
TGGTCTGCGGTGGCTGGCAGACCAAGGCTGGGTGCCCTCCCGCGGCTGTGCGCGTATTGGGGCTCT
GACTGGGCTGGGCTGTGGGTGGTGGGCTGGTGGCGGCGAAGAGAGCTGGGGGGTCCC
CTGCTGGCGCGGCTGTGGCCTATGGGCTGAGCGCGGGGAGTTACGCCCCGCTGGTTTTCGGTGTACTCC
CCGGGCTGGTGGGCGTCCGAGGTGTGGTGCAGGCCACAGGCTGGTATGATGCTGATGAGCCTCGGGGG
GCTCCTGGGCCCTCCCTGTCAAGTAAGTTCTTAAGGATGAGACAGGAGACTTACCGCCTCTTTCTC
CTGTCTGGTTCTTTGATCCTCTCCGCGAGCTTATCTACATAGGGTTGCCAGGCGCTGCCCTCCTGTG
GTCCAGCCTCCCTCCAGCCACGCTCCCCAGAGACGGGGAGCTGCTTCCCGCTCCCGAGCAGTCTT
GCTGTCCCCAGAGGCCCTGGCTCCACTCTGGACACCACTTGTGATATTATTTCTGTTTGGAGCCCTCC
CCCAATAAAGAAATTTTATCGGGTTTTCCTGAAACCTCCAAC

The disclosed NOV5d nucleic acid sequence, located on chromosome 17, has 567 of 940 bases (60%) identical to a gb:GENBANK-ID:HSU81800|acc:U81800.1 mRNA from *Homo sapiens* (monocarboxylate transporter (MCT3) mRNA, complete cds) ($E = 6.5e^{-30}$).

A disclosed NOV5d polypeptide (SEQ ID NO:30) encoded by SEQ ID NO:29 is 472 amino acid residues and is presented using the one-letter amino acid code in Table 5H. Signal P, Psort and/or Hydropathy results predict that NOV5d contains no signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.6000. Alternatively, NOV5d is also likely to be localized to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a certainty of 0.3000

Table 5H. Encoded NOV5d protein sequence (SEQ ID NO:30).

MPAPQRKHRRGGFSHRCFPTPTAMTPQAPGPPDGGWGVVAAAAFAINGLSYGLLRSLGLAFPDIAEHFDRSAQ
DTAWISALALAVQQAASPVGSALSTRWGARPVVMVGVLASLGFVFSAFASDLLHLHLGLLAGFGWALVFAPA
LGTLSTYFSRRRVLAVGLALTNGASSLLAPALQLLDTFGWGALLLLGATTLHLTPCGALLPLVLPDPDPA
PPRSPLAALGLSLFTRRAFSIFALGTALVGGGYFVPYVHLAPHALDRGLGGYGAALVVAVAAAMDAGARLVCGWL
ADQGWVPLPRLLAVFGALTGLGLVWVGLVPVVGGEESWGGPLLAAVAYGLSAGSYAPLVFVGLPGLVGVGGVQ
ATGLVMMLMSLGGLLGPPLSGFLRDETGFDTASFLLSGSLILSGSFIYIGLPRALPSCGPASPPATPPPETGEL
LPAPQAVLLSPGGPGSTLDTTC

The disclosed NOV5d amino acid sequence has 96 of 198 amino acid residues (48%) identical to, and 122 of 198 amino acid residues (61%) similar to, the 504 amino acid residue

ptnr:SPTREMBL-ACC:O95907 protein from *Homo sapiens* (Human) (DJ1039K5.2 (Similar To Monocarboxylate Transporter (MCT3))) ($E = 7.9e^{-68}$).

NOV5d is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. .

NOV5e

A disclosed NOV5e nucleic acid of 465 nucleotides (also referred to as CG56635-05) encoding a novel Monocarboxylate transporter 3-like protein is shown in Table 5I. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 7-9 and ending with a TGA codon at nucleotides 436-438. The start and stop codons are in bold letters in Table 5I., and the 5' and 3' untranslated regions, if any, are underlined.

Table 5I. NOV5e Nucleotide Sequence (SEQ ID NO:31)

ACGGCGATGACCCCCAGCCCCGCGACCCCCGGATGGGGGCTGGGGCTGGGTGGTGGCGGCCGAGCCT
TCGCGATAAACGGGCTGTCCTACGGGCTGCTGCGCTCGCTGGGCCTTGCCTTCCCTGTCCTTCCCGAGCA
CTTTGACCGAAGCGCCAGGACACTGCGTGGATCAGCGCCCTGGCCCTGGCCGTGCAGCAGGCAGCCAGC
TTCTAAGGGATGAGACAGGAGACTTCACCGCCTCTTTCCTCCTGTCTGGTTCTTTGATCCTCTCCGGCA
GCTTCATCTACATAGGGTTGCCAGGGCGCTGCCCTCCTGTGGTCCAGCCTCCCCCTCCAGCCACGCCTCC
CCCAGAGACGGGGAGCTGCTTCCCGCTCCCCAGGCAGTCTTGCTGTCCCCAGGAGGCCCTGGCTCCACT
CTGGACACCACTTGT**TGATTATTTCT**GTTTGAGCCCTCCCC

The disclosed NOV5e nucleic acid sequence, located on chromosome 17, has 351 of 434 bases (80%) identical to a gb:GENBANK-ID:AX083362|acc:AX083362.1 mRNA from *Homo sapiens* (Sequence 54 from Patent WO0112660) ($E = 1.6e^{-53}$).

A disclosed NOV5e polypeptide (SEQ ID NO:32) encoded by SEQ ID NO:31 is 143 amino acid residues and is presented using the one-letter amino acid code in Table 5J. Signal P, Psort and/or Hydropathy results predict that NOV5e contains no signal peptide and is likely to be localized extracellularly with a certainty of 0.5040. Alternatively, NOV5e is also likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.1000, to the endoplasmic reticulum (lumen) with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV5e is between positions 43 and 44: VLA-EH.

Table 5J. Encoded NOV5e protein sequence (SEQ ID NO:32).

MTPQPAGPPDGGWGVVAAAAFAINGLSYGLLRSLGLAFPVLAEHFDRSAQDTAWISALALAVQQAASFLRDETG
DFTASFLSLSGSLILSGSFIYIGLPRALPSCGPASPATPPPETGELLPAQAVLLSPGGPGSTLDTTC

The disclosed NOV5e amino acid sequence has 67 of 68 amino acid residues (98%) identical to, and 67 of 68 amino acid residues (98%) similar to, the 375 amino acid residue ptnr:REMTREMBL-ACC:CAC33285 protein from *Homo sapiens* (Human) (Sequence 54 from Patent WO0112660) ($E = 2.9e^{-31}$).

NOV5e is predicted to be expressed in at least Mammalian Tissue, Parathyroid Gland, Mammary gland/Breast, Prostate. .

NOV5a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 5K.

Table 5K. BLAST results for NOV5a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7670446 dbj BAA95074.1 (AB041591)	unnamed protein product [<i>Mus musculus</i>]	290	252/288 (87%)	263/288 (90%)	1e-86
gi 17491104 ref XP_064368.1 (XM_064368)	similar to solute carrier family 16 (monocarboxylic acid transporters), member 8 (H. sapiens) [<i>Homo sapiens</i>]	427	196/398 (49%)	257/398 (64%)	6e-74
gi 2497855 sp Q63344 MOT2_RAT	MONOCARBOXYLATE TRANSPORTER 2 (MCT 2)	489	142/420 (33%)	220/420 (51%)	6e-53
gi 1432167 gb AAB04023.1 (U62316)	monocarboxylate transporter 2 [<i>Rattus norvegicus</i>]	489	143/420 (34%)	220/420 (52%)	6e-53
gi 6755536 ref NP_035521.1 (NM_011391)	solute carrier family 16 (monocarboxylic acid transporters), member 7 [<i>Mus musculus</i>]	484	142/421 (33%)	221/421 (51%)	2e-52

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 5J.

Table 5J Information for the ClustalW proteins

- 1) NOV5a (SEQ ID NO:24)
- 2) NOV5b (SEQ ID NO:26)
- 3) NOV5c (SEQ ID NO:28)
- 4) NOV5d (SEQ ID NO:30)
- 5) NOV5e (SEQ ID NO:32)
- 6) gi|7670446|dbj|BAA95074.1| (AB041591) unnamed protein product [*Mus musculus*] (SEQ ID NO:337)

[illegible]

			610	620	630	640	650	660	
5	NOV5a	426	LPAPQAVLLSPG	-----	GPSTLDTTC	-----	-----	-----	447
	NOV5b	170	LPAPQAVLLSPG	-----	GPSTLDTTC	-----	-----	-----	191
	NOV5c	194	LPAPQAVLLSPG	-----	GPSTLDTTC	-----	-----	-----	215
	NOV5d	451	LPAPQAVLLSPG	-----	GPSTLDTTC	-----	-----	-----	472
	NOV5e	122	LPAPQAVLLSPG	-----	GPSTLDTTC	-----	-----	-----	143
10	gi 7670446	235	LAAAGAYGLSAGSYAPLVFGVLPGLVGIGGVVQATGLVMMLMSLGGLLGPPLSGKG	----	-----	-----	-----	-----	290
	gi 17491104	309	VALAVAYGFTSCALAPLAFSVLPGLIGTRRIYCGLGLLQMIESIGGLLGPPLSGYL RDVT	----	-----	-----	-----	-----	368
	gi 2497855	345	MVYVIFFGIGFGSISLLFECLMDQVGASRFSSAVGLVTIVECCPVLFPGPPLAGKLLDIT	----	-----	-----	-----	-----	404
	gi 1432167	345	MVYVIFFGIGFGSISLLFECLMDQVGASRFSSAVGLVTIVECCPVLFPGPPLAGKLLDIT	----	-----	-----	-----	-----	404
	gi 6755536	340	MVYVVFGLGFGSISLLFECLMDIVCATRFSSAVGLTIVECCPVLFPGPPLAGKLLDIT	----	-----	-----	-----	-----	399
15			670	680	690	700	710	720	
	NOV5a	447	-----	-----	-----	-----	-----	-----	447
	NOV5b	191	-----	-----	-----	-----	-----	-----	191
	NOV5c	215	-----	-----	-----	-----	-----	-----	215
20	NOV5d	472	-----	-----	-----	-----	-----	-----	472
	NOV5e	143	-----	-----	-----	-----	-----	-----	143
	gi 7670446	290	-----	-----	-----	-----	-----	-----	290
	gi 17491104	369	GNYTASFVVGAFLLSGSGILL'LPHFFC	-----	-----	-----	-----	FSTTTTSG	404
	gi 2497855	405	GQYKLYIASGIVVLSSGIYLLICNAINYRLLEKERKREKARRKKSASQASKEMEALSRS	----	-----	-----	-----	-----	464
25	gi 1432167	405	GQYKLYIASGIVVLSSGIYLLICNAINYRLLEKERKREKARRKKSASQASKEMEALSRS	----	-----	-----	-----	-----	464
	gi 6755536	400	GEYKLYIASGTVVLVSGTYLLIGNAINYRLLDKERKREKAKKKKSASHASREMEALNRS	----	-----	-----	-----	-----	459
30			730	740					
	NOV5a	447	-----	-----	-----	-----	-----	-----	447
	NOV5b	191	-----	-----	-----	-----	-----	-----	191
	NOV5c	215	-----	-----	-----	-----	-----	-----	215
	NOV5d	472	-----	-----	-----	-----	-----	-----	472
	NOV5e	143	-----	-----	-----	-----	-----	-----	143
35	gi 7670446	290	-----	-----	-----	-----	-----	-----	290
	gi 17491104	405	PQDLVTEALDTKVPLPKEGLEED	--	427	-----	-----	-----	427
	gi 2497855	465	KQDDVTVKVSNTHNPPSDRDKESSI	489	-----	-----	-----	-----	489
	gi 1432167	465	KQDDVTVKVSNTHNPPSDRDKESSI	489	-----	-----	-----	-----	489
40	gi 6755536	460	KQDEVTVKASNAHNPPSDRDKESNI	484	-----	-----	-----	-----	484

Monocarboxylates such as lactate and pyruvate play a central role in cellular metabolism and metabolic communication between tissues. Essential to these roles is their rapid transport across the plasma membrane, which is catalysed by a recently identified family of proton-linked monocarboxylate transporters (MCTs). Nine MCT-related sequences have so far been identified in mammals, each having a different tissue distribution, whereas six related proteins can be recognized in *Caenorhabditis elegans* and 4 in *Saccharomyces cerevisiae*. Direct demonstration of proton-linked lactate and pyruvate transport has been demonstrated for mammalian MCT1-MCT4, but only for MCT1 and MCT2 have detailed analyses of substrate and inhibitor kinetics been described following heterologous expression in *Xenopus* oocytes. MCT1 is ubiquitously expressed, but is especially prominent in heart and red muscle, where it is up-regulated in response to increased work, suggesting a special role in lactic acid oxidation. By contrast, MCT4 is most evident in white muscle and other cells with a high glycolytic rate, such as tumour cells and white blood cells, suggesting it is expressed where lactic acid efflux predominates. MCT2 has a ten-fold higher affinity for substrates than MCT1

and MCT4 and is found in cells where rapid uptake at low substrate concentrations may be required, including the proximal kidney tubules, neurons and sperm tails. MCT3 is uniquely expressed in the retinal pigment epithelium. MCT1 and MCT4 have been shown to interact specifically with OX-47 (CD147), a member of the immunoglobulin superfamily with a single transmembrane helix. This interaction appears to assist MCT expression at the cell surface

The disclosed NOV5 nucleic acid of the invention encoding a Monocarboxylate transporter (MCT3)-like protein includes the nucleic acid whose sequence is provided in Table 5A, 5C, 5E, 5G, 5I or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 5A, 5C, 5E, 5G, or 5I while still encoding a protein that maintains its Monocarboxylate transporter (MCT3)-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 40 percent of the bases may be so changed.

The disclosed NOV5 protein of the invention includes the Monocarboxylate transporter (MCT3)-like protein whose sequence is provided in Table 5B, 5D, 5F, 5H, or 5J. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 5B, 5D, 5F, 5H, or 5J while still encoding a protein that maintains its Monocarboxylate transporter (MCT3)-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 67 percent of the residues may be so changed.

NOV5 nucleic acid and polypeptide show homology to the Monocarboxylate transporter (MCT3) family of proteins. Accordingly, to the NOV5 nucleic acid and polypeptide may function as members of this family. The NOV5 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this

invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The nucleic acids and proteins of NOV5 are useful in metabolic disorders such as salla disease, infantile sialic acid storage disease, symptomatic deficiency in lactate transport, subnormal erythrocyte lactate transport, muscle injuries, cystinosis, streptozotocin-induced diabetes, hypoxia, cardiac arrest or stroke, neuronal disorders, retinal angiogenesis, and/or other pathologies and disorders.

NOV5 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV5 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV6

A disclosed NOV6 nucleic acid of 1336 nucleotides (also referred to CG56674-01) encoding a novel Nitrilase-1-like protein is shown in Table 6A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 77-79 and ending with a TAA codon at nucleotides 1058-1060. In Table 6A, the 5' and 3' untranslated regions are underlined and the start and stop codons are in bold letters.

Table 6A. NOV6 Nucleotide Sequence (SEQ ID NO:33)

GCCCACTCGCTGCGGCCTATCTGGGCTCCAGACCGCCCTCCGGATCGGACCCTGCGAATGGTTTTGGCTATA
TCTTCA**TG**CTGGGCTTCATCACCAGGCGCTCTCAGAGATTCCCTGTCCTTCTGTGTCTGGACTCCGGATA
CCTCAACTCTCTGGGGAAGGTGCTCAGCCAGGCCAGACCAATGGCTATCTCCTCTCTCTCGGA
GCCCTGGTGGCTGTGTGCGAGTAAACATCGACGCAGACAGAACAAGAACTTTAAACATGTGCTGAG
TGGTTCGAGAGGCTGCAGACTGGGTGCCGTCCTGGCTTTCTGCTGAGGCATTGACTTCATTGCACGG
GACCTGCGAGAGACGCTACACCTGTCTGAACCACTGGGTGGGAACTTTTGAAGAATAACACCGACTTGC
CAGGGAATGTGGACTCTGGCTTCTCTGGGTGGTTTCCATGAGCTGGCCAAAGACTGGGAGCAGACTCAGA
AAATCTACAA**T**TGTACGTGCTGCTGAACAGCAAAGGGGCAGTAGTGGCCATTACAGGAAGACACATCTG
TGTGACGTAGAGATTCACGGGCAGGGGCCATGTGTGAAAGCAACTCTACCATGCCTGGGCCAGTCTTGA
GTACCTTCTCAGCACACCCAGCAGCAAGATTGGTCTAGCTGTCTGCTATGACATGCGGTTCCCTGA
ACTCTGGCATGGCTCAAGCTGGAACAGAGATACTTACTTACTCTCAGCTTTTGGATCATTACAGGCCCA
GCCCACTGGGAGGTGTTGTCTGCGGGCCGTGCTATCGAAACCCAGTGCTATGTAGTGGCAGCAGCACAGTG
TGGACGCCACCATGAGAAGAGAGCAAGTTATGGCCACAGCATGTGTGGTAGACCCCTGGGGGAACAGTGGTG
CCCGCTGCTCTGAGGGGCCAGGCCCTTGCTTGGCCGAATAGACCTCAACTATCTGCGACAGTTCGCGGA
CACCTGCCTGTGTTTCCAGCACCGCAGGCCTGACCTTATGGCAATCTGGGTACCCCACTGTCT**TA**GACTT
GACTTCTGTGAGTTTAGACCTGCCCCCTCCACCCCCACCCTGCCACTATGAGCTAGTGCTCATGTGACTTG
GAGCGAGGATCCAGGCACAGTCCCCCTCACTGGGAACCTTGACTCTTTGATGGAACACAGATGGGCTG
CTTGGGAAGA**AA**ACTTTCACTTGAGCTTCACTTGAGTGAGTCAAGTCAAGTTTCAAGAAAGTGAATTTATA
TAGTCA**T**TGTTTATTTCATGAAACTGAAGTTCTGCTGAGGGCTGAGCACCTTCCCCA

The disclosed NOV6 nucleic acid sequence, localized to the p14.2 region of chromosome 3, has 1319 of 1329 bases (99%) identical to a gb:GENBANK-ID:AF069987|acc:AF069987.1 mRNA from *Homo sapiens* (nitrilase 1 (NIT1) mRNA, complete cds) ($E = 3.1 \times 10^{-290}$).

A disclosed NOV6 polypeptide (SEQ ID NO:34) encoded by SEQ ID NO:33 is 327 amino acid residues and is presented using the one-letter amino acid code in Table 6B. Signal P, Psort and/or Hydropathy results predict that NOV6 has a signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.4500. Alternatively, NOV6 is also likely to be localized to the microbody (peroxisome) with a certainty of 0.3000, to the lysosome (lumen) with a certainty of 0.2021, or to the mitochondrial matrix space with a certainty of 0.1000. The most likely cleavage site for NOV6 is between positions 27 and 28: LSG-EG

Table 6B. Encoded NOV6 protein sequence (SEQ ID NO:34).

MLGFI^{TR}PPH^RFLS^LLLCPGLRIPQLSGEGAQPRPRAMAISSSSCELPVAVCQVTSTPDKQONFKTCAELV
REAA^RLGA^CLAFLPEAFDFIARDPAETLHLSEPLGGKLEEY^TQLARECGLWLSLGGFHERGQDWEQTQKI
YNCHVLNSKGAVVAIYRK^THLCDVEIPGGQPMCESNTMPGSPLES^PVSTPAKGIGLAVCYDMR^PPELS
ALAAGCEILTYPSAFGSI^TGPAHWEVLLRAAIETQCYVVA^AAQCRHHEKRASYGHSVMVDPWGT^VVAR
CSEGPGLCLARIDLNYLRQLRRLHPV^FQHRRPDLYGNLGHPLS

15 The disclosed NOV6 amino acid sequence has 322 of 327 amino acid residues (98%) identical to, and 322 of 327 amino acid residues (98%) similar to, the 327 amino acid residue ptnr:SPTREMBL-ACC:O76091 protein from *Homo sapiens* (Human) (Nitrilase Homolog 1) ($E = 4.5e^{-176}$).

NOV6 also has homology to the amino acid sequence shown in the BLASTP data listed in Table 6C.

Table 6C. BLAST results for NOV6					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 5031947 ref NP_005591.1 (NM_005600)	nitrilase 1 [<i>Homo sapiens</i>]	327	322/327 (98%)	322/327 (98%)	0.0
gi 3242980 gb AAC40184.1 (AF069985)	nitrilase homolog 1 [<i>Mus musculus</i>]	323	272/327 (83%)	298/327 (90%)	e-154
gi 6754856 ref NP_036179.1 (NM_012049)	nitrilase 1 [<i>Mus musculus</i>]	323	272/327 (83%)	297/327 (90%)	e-153
gi 18204913 gb AAH21634.1 AAH21634 (BC021634)	Unknown (protein for MGC:13825) [<i>Mus musculus</i>]	323	271/327 (82%)	296/327 (89%)	e-153
gi 12836591 dbj BAB23723.1 (AK004988)	data source:MGD, source key:MGI:1350916, evidence:ISS-nitrilase 1-putative [<i>Mus musculus</i>]	290	251/288 (87%)	272/288 (94%)	e-145

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 6D.

Table 6D. Information for the ClustalW proteins

- 1) NOV6 (SEQ ID NO:34)
- 2) gi|5031947|ref|NP_005591.1| (NM_005600) nitrilase 1 [*Homo sapiens*] (SEQ ID NO:342)
- 3) gi|3242980|gb|AAC40184.1| (AF069985) nitrilase homolog 1 [*Mus musculus*] (SEQ ID NO:343)
- 4) gi|6754856|ref|NP_036179.1| (NM_012049) nitrilase 1 [*Mus musculus*] (SEQ ID NO:344)
- 5) gi|18204913|gb|AAH21634.1|AAH21634 (BC021634) Unknown (protein for MGC:13825) [*Mus musculus*] (SEQ ID NO:345)
- 6) gi|12836591|dbj|BAB23723.1| (AK004988) data source:MGD, source key:MGI:1350916, evidence:ISS-nitrilase 1-putative [*Mus musculus*] (SEQ ID NO:346)

			10	20	30	40	50	60		
NOV6	1	MLGFITRPPH	RFLS	LLCPGLRIP	OLSGEGA	QPRPRAM	AISSSS	CELPLVAVCQVTSTPDK	60	
gi 5031947	1	MLGFITRPPH	RFLS	LLCPGLRIP	OLSVLCA	QPRPRAM	AISSSS	CELPLVAVCQVTSTPDK	60	
gi 3242980	1	MLGFITRPPH	--	QLLCTGYRLL	RIPVLCT	QPRPRIMS	--	SSTSWELPLVAVCQVTSTPNK	56	
gi 6754856	1	MLGFITRPPH	--	QLLCTGYRLL	RTPVLCT	QPRPRIMS	--	SSTSWELPLVAVCQVTSTPNK	56	
gi 18204913	1	MLGFITRPPH	--	QLLCTGYRLL	RTPVLCT	QPRPRIMS	--	SSTSWELPLVAVCQVTSTPNK	56	
gi 12836591	1	-----	-----	-----	-----	MS	-----	SSTSWELPLVAVCQVTSTPNK	23	
			70	80	90	100	110	120		
NOV6	61	QENFKTCAELV	REAA	RLGAC	LAFL	PEAF	DFI	ARDPA	ETLHLSEPLGCKLLEBYTQLAREC	120
gi 5031947	61	QENFKTCAELV	REAA	RLGAC	LAFL	PEAF	DFI	ARDPA	ETLHLSEPLGCKLLEBYTQLAREC	120
gi 3242980	57	QENFKTCAELV	QEAAR	LGAC	LAFL	PEAF	DFI	ARNPA	ETLLSEPLNGDLLGQYSQLAREC	116
gi 6754856	57	QENFKTCAELV	QEAAR	LGAC	LAFL	PEAF	DFI	ARNPA	ETLLSEPLNGDLLGQYSQLAREC	116
gi 18204913	57	QENFKTCAELV	QEAAR	LGAC	LAFL	PEAF	DFI	ARNPA	ETLLSEPLNGDLLGQYSQLAREC	116
gi 12836591	24	QENFKTCAELV	QEAAR	LGAC	LAFL	PEAF	DFI	ARNPA	ETLLSEPLNGDLLGQYSQLAREC	83
			130	140	150	160	170	180		
NOV6	121	GIWLSLGGF	FHERG	QDWE	QTKI	YNCH	VLLNS	KGAV	VATYRKTHLCDVEIPGQGPMCESNS	180
gi 5031947	121	GIWLSLGGF	FHERG	QDWE	QTKI	YNCH	VLLNS	KGAV	VATYRKTHLCDVEIPGQGPMCESNS	180
gi 3242980	117	GIWLSLGGF	FHERG	QDWE	QNQKI	YNCH	VLLNS	KGSV	VASYRKTHLCDVEIPGQGPMRESNY	176
gi 6754856	117	GIWLSLGGF	FHERG	QDWE	QNQKI	YNCH	VLLNS	KGSV	VASYRKTHLCDVEIPGQGPMRESNY	176

	gi 18204913	117	GIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHLCDVEIPGQGPMRESNY	176
	gi 12836591	84	GIWLSLGGFHERGQDWEQNQKIYNCHVLLNSKGSVVASYRKTHLCDVEIPGQGPMRESNY	143
5			190 200 210 220 230 240	
	NOV6	181	TMPGPSLESFVSTPAGKIGLAVCYDMRFPELSLALAQAGTEILTYPYSAFGSITGPAHWEV	240
	gi 5031947	181	TMPGPSLESFVSTPAGKIGLAVCYDMRFPELSLALAQAGAEILTYPYSAFGSITGPAHWEV	240
	gi 3242980	177	TKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAAQAGAEILTYPYSAFGSVTGP	236
	gi 6754856	177	TKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAAQAGAEILTYPYSAFGSVTGP	236
10	gi 18204913	177	TKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAAQAGAEILTYPYSAFGSVTGP	236
	gi 12836591	144	TKPGGTLEPPVKTPAGKVGLAICYDMRFPELSLKLAAQAGAEILTYPYSAFGSVTGP	203
			250 260 270 280 290 300	
15	NOV6	241	LLRARAIEIQCYVVAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLNY	300
	gi 5031947	241	LLRARAIEIQCYVVAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLNY	300
	gi 3242980	237	LLRARAIESQCYVIAAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLHF	296
	gi 6754856	237	LLRARAIESQCYVIAAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLHF	296
	gi 18204913	237	LLRARAIESQCYVIAAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLHF	296
20	gi 12836591	204	LLRARAIESQCYVIAAQAQCGRHHEKRASYGHSMVVDPWGTVVARCSEGPGLCLARIDLHF	263
			310 320	
	NOV6	301	LQQLRRHLPVVFQHRRPDLYGSLGHPLS	327
25	gi 5031947	301	LQQLRRHLPVVFQHRRPDLYGSLGHPLS	327
	gi 3242980	297	LQQMRQHLPVVFQHRRPDLYGSLGHPLS	323
	gi 6754856	297	LQQMRQHLPVVFQHRRPDLYGSLGHPLS	323
	gi 18204913	297	LQQMRQHLPVVFQHRRPDLYGSLGHPLS	323
30	gi 12836591	264	LQQMRQHLPVVFQHRRPDLYGSLGHPLS	290

Tables 6E list the domain description from DOMAIN analysis results against NOV6. This indicates that the NOV6 sequence has properties similar to those of other proteins known to contain this domain.

Table 6E. Domain Analysis of NOV6

gnl|Pfam|pfam00795, CN_hydrolase, Carbon-nitrogen hydrolase. This family contains hydrolases that break carbon-nitrogen bonds. The family includes: Nitrilase EC:3.5.5.1, Aliphatic amidase EC:3.5.1.4, Biotidinase EC:3.5.1.12, Beta-ureidopropionase EC:3.5.1.6. (SEQ ID NO:803)
CD-Length = 267 residues, 100.0% aligned
Score = 273 bits (698), Expect = 1e-74

	NOV 6:	51	VCQVTSTP-DKQQNFKTCAELVREAAALGACLAFLPEAFDFI---ARDPAETLHLSEPLG	106
	Sbjct:	1	AVQAEPVPEDLAANLQKAEELIEEAAGAEELVVFPEAFIPGYPYCKSDAEYYENAEAD	60
40				
	NOV 6:	107	GKLLLEYTQLARECGLWLSLGGFHERGQDWEQTQKIYNCHVLLNSKGAVVAIYRKTHLCD	166
	Sbjct:	61	GEETQFLSRLARKNGIVIVLGVSEREGEG-----KLYNTAVLIDPDGKLIKGYRKIHLFT	115
45				
	NOV 6:	167	V---EIPGQGPMECSNSTMPGPSLESFVSTPAGKIGLAVCYDMRFPELSLALAQAGTEIL	223
	Sbjct:	116	DPERKVYGE-----GGSGFPVFDTPVGKLGLLICYDIRFPELARALALKGAEL	165
50				
	NOV 6:	224	TYPYSAFGSITGPAHWEVLLRARAIEIQCYVVAQAQCGRHHEKRA-----SYGHSMVVDPW	278
	Sbjct:	166	AWPSAFGRKTGDSHWELLARARAIEINQCFVAAANQVGTEEDLDFDLGFEFYGHSMIIDPD	225
	NOV 6:	279	GTVVA-RCSEGPGLCLARIDLNLRLQLRRHLPVVFQHRRPDLY	319

6B while still encoding a protein that maintains its Nitrilase-1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18 percent of the residues may be so changed.

The protein homology information, expression pattern, and map location for the Nitrilase-1-like protein and nucleic acid (NOV6) disclosed herein suggest that NOV6 may have important structural and/or physiological functions characteristic of the Nitrilase-1-like family. Therefore, the NOV6 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo.

The NOV6 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, muscle conditions, disorders and diseases, longevity, and/or other pathologies/disorders. The NOV6 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

NOV6 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV6 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV7

NOV7 includes three novel cleavage signal-1 protein-like proteins disclosed below. The disclosed sequences have been named NOV7a, NOV7b, NOV7c, and NOV7d.

NOV7a

5 A disclosed NOV7a nucleic acid of 1822 nucleotides (also referred to as CG56613-01) encoding a novel cleavage signal-1 protein-like protein is shown in Table 7A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 98-100 and ending with a TAA codon at nucleotides 839-841. A putative untranslated region upstream from the initiation codon is underlined in Table 7A. The start and stop codons are in
10 bold letters.

Table 7A. NOV7a nucleotide sequence (SEQ ID NO:35).

<p>GGGGCTGACGCAGCATTGCCAATTCTAAATCCATCATTGACTGAGGAGGAGAGGTTGAAGTTGATCAGCT CCAGGGTTTGAGAAATTCAGTCCGAATGGAACCTCAGGACCTGGAACCTGCAGCTGGAGGAGCGCTGCTGGG CCTGGAGGAGCAGCTTCGTGCTGTGCGCATGCCTTCACCCCTCCGCTCCTCCGCACTCATGGGAATGTGTGG CAGTAGAAGCACTGATAACTTGTCTATGCCCTTCTCCATTGAATGTAATGGAACCACTGACTGAAGTATGCA GGAGCAGTCATACCTGAAGTCTGAATTGGGCTGGGACTTGGAGAAATGGGATTGAAATTCCTCCTGGAGA AAGCTCAGAATCTGTTTTTCCAAGCAACGATCAGAATCATCTTCTATATGTTCTGGTCCCTCTCATGCTAA CAGAAGAACTGGAGTACCTTCTACTGCTCAGTGGGCAAAATCCAAAACCCCATTTAGTGGCAAGGAAGAAAGT GTTCCGAGCATCGGTGGCTCTAACGCCAACAGCTCCTTCTAGAACAGGCTCTGTGCAGACACCTCCAGATTT GGAAAGTTCTGAGGAAGTTGATGCAGCTGAAGGAGCCCCAGAAGTTGAGGACCTAAATCTGAAGTGGGAAGA AGGGCATGGAAAACCTCCATCAATGCCAGCTGCTGAGGAAATGCATAAAAATGTGGAGCAAGATGAGTTGCA GCAAGTCATACGGGAGATTAAAGAGTCTATTGTTGGGGAAATCAGACGGGAAATGTAAGTGGACTTTTGGC AGCAGTATCTCAAGTAAAGCGTCTAATTCTAAGCAAGATTATCATTAAACAGAAATATAGGTTGGCATGG ATCCTATTAGCTGTGTAATACTGGAATTATCAATGATATGCACTGGTGGAGGTGTTATTTGTGCTTTAGAAG ATACTTGCTGTTGAGCTGGGCTACTGTATACAGTGTACAAATGTGTATTTCTTCAACCATATATTTTAAAAAG ACGTACATAGAACTTAGGCACCTTGTCTATTTCTTTCTAACTATCAAAAACCTTAGCAGTTTGAAAAGCC TAATATTTATTTGTATGTCAATATTTTTCATTGATTCCCTATTAGAATTAATTTTAAAACTTGAAGACTTC CAGACTTATCCAACTTATAAATAACATATTTCTTCAGACTAACATCTTAAAAACACTGACCTCTATGAGGTAT TTACTGTGCAATAACTGATTCAATTTTTTTCAGAGCTTGAAGCATCCAATGATTTTTTCCCTCCACTGCTGTTA ATTAATGTCACTTCCAAGAAGAAAACTGTTCTGTTGTAAAAAATATAATTGCTCTTAATCTTGGGGAGGT TACTAATAGCAGTAGGATAGAATTTTATGAGGTTACCTACAACACTTAATGTACTTACACTGTAAGCCTTG TTGCTTTACCCAAGACAAATGTAATTTTATCATTTGCTTATGTAGTATTTTCTTTGGAAATGTGCCTTATG TTAAACACTATGTACTTTTACTTTTTGCATTGTCCAGACTTCTTTATTAGATGGAGATGTTTCTTTTCTGT CTTCTAGACTAAATAGAGTATCATCAAATAATGGGGCCTATGACTTGAATGAATAGAAATGAATAAGCTGG TGTTTGTTTTTTCAAAATGGAAGTAATTTAGATTTGTTCTCCTCATACATAAAATGATTTTAGTTTCAGTTT AACCAGTGAAAACCTTTGTTTTATGAAAAAAGGAAAATGGTTTCCATTGTTTATATGTGTTAAATA AATGTGTAAAGTAACCACCCC</p>

15 The disclosed NOV7a nucleic acid sequence, localized to chromosome 2, has 1822 of 1828 bases (99%) identical to a gb:GENBANK-ID:HUMCS1PA|acc:M61199.1 mRNA from *Homo sapiens* (Human cleavage signal 1 protein mRNA, complete cds) (E = 0.0).

The disclosed NOV7a polypeptide (SEQ ID NO:36) encoded by SEQ ID NO:35 has 247 amino acid residues and is presented in Table 7B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV7a has a signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500. Alternatively, NOV7A may also

localize to the mitochondrial matrix space with a certainty of 0.1000, or the lysosome (lumen) with a certainty of 0.1000.

Table 7B. Encoded NOV7a protein sequence (SEQ ID NO:36).

MELQDLELQLEERLLGLEEQLRAVRMPSPPFRSSALMGMCGRSTDNLSCPSPLNVMEPVTELMQEESYLKSE LGLGLGEMGFEPGESSSEVFSKQRSESSSICSGPSHANRRRTGVPSTASVGKSKTPLVARKKVFRASVALT PTAPSRGTGSVQTTPDLESSEEVDAEAGAPEVVGPKSEVEEGHGKLPMPAAEEMHKNVEQDELQQVIREIKE SIVGEIRREIVSGLLAAVSSSKASNSKQDYH
--

A search of sequence databases reveals that the NOV7a amino acid sequence has 247 of 249 amino acid residues (99%) identical to, and 247 of 249 amino acid residues (99%) similar to, the 249 amino acid residue ptnr:SWISSPROT-ACC:P28290 protein from *Homo sapiens* (Human) (Sperm-Specific Antigen 2 (Cleavage Signal-1 Protein) (CS-1)) ($E = 6.1e^{-124}$).

NOV7a is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Aorta, Ascending Colon, Bone, Cervix, Cochlea, Colon, Dermis, Gall Bladder, Hypothalamus, Islets of Langerhans, Liver, Lung, Lymphoid tissue, Ovary, Parathyroid Gland, Parotid Salivary glands, Pineal Gland, Retina, Right Cerebellum, Skin, Tonsils, Umbilical Vein, Vein, Whole Organism.

NOV7b

In the present invention, the target sequence identified previously, NOV7a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in

PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated Accession Number NOV7b (6 aminoacid different from NOV7a) and NOV7c (2 aminoacid different from NOV7a).

A disclosed NOV7b nucleic acid of 806 nucleotides (also referred to as CG56613-02) encoding a novel cleavage signal-1 protein-like protein is shown in Table 7C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 21-23 and ending with a TAA codon at nucleotides 762-764. A putative untranslated region upstream from the initiation codon is underlined in Table 7C. The start and stop codons are in bold letters, and the 5' and 3' untranslated regions, if any, are underlined.

Table 7C. NOV7b nucleotide sequence (SEQ ID NO:37).

GTTTGAGAAATTCAGTCCGAATGGAACTTCAGGACCTGGAAGTGCAGCTGGAGGAGCGCCTGCTGGGCCTGG
AGGAGCAGCTTCGTGCTGTGCGCATGCCTTCACCTTCCGCTCCTCCGCACTCATGGGAATGTGTGGCAGTA
GAAGCGCTGATAACTTGTTCATGCCCTTCTCCATTGAATGTAATGGAACCACTCACTGAACCTGATGCAGGAGC
AGTCATACCTGAAGTCTGAATTGGGCCTGGGACTTGGAGAAATGGGATTGAAATTCCTCCTGGAGAAAGCT
CAGAATCTGTTTTTCCCAAGCAACATCAGAATCATCTTCTGTATGTTCTGGTCCCTCTCATGTCTAACAGAA
GAAGTGGAGTACCTTCTACTGTCTCAGTGGGCAATCCAAACCCCATAGTGGCAAGGAAGAAAGTGTTC
GAGCATCGGTGGCTCTAACGCCAACAGCTCCTTCTAGAACAGGCTCTGTGCAGACACCTCCAGATTTGGAAA
GTTCTGAGGAAGTTGATGCAGCTGAAGGAGCCCCAGAAGTTGTAGGACCTAAATCTGAAGTGGGAAGAAGGGC
ATGGAAAACTCCCATCAATGCCAGCTGTTGAGGAAATGCATAAAATGTGGAGCAAGATGAGTTGCAGCAAG
TCATACGGGAGATTAAAGAGTCTATTGTTGGGGAAATCAGACGGGAAATGTAAAGTGGACTTTTGGCAGCAG
TATCTTCAAGTAAAGCGTCTAATTCTAAGCAAGATTATCATTAACAGAAATTATAGGTTGGCATGGATCCT
ATTAGCTGTGTAAT

The disclosed NOV7b nucleic acid sequence, localized to chromosome 2, has 801 of 812 bases (98%) identical to a gb:GENBANK-ID:HUMCS1PA|acc:M61199.1 mRNA from *Homo sapiens* (Human cleavage signal 1 protein mRNA, complete cds) ($E = 7.6e^{-171}$).

The disclosed NOV7b polypeptide (SEQ ID NO:38) encoded by SEQ ID NO:37 has 247 amino acid residues and is presented in Table 7D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV7b has no signal peptide and is

likely to be localized to the cytoplasm with a certainty of 0.6500. Alternatively, NOV7b may also localize to the mitochondrial matrix space with a certainty of 0.1000, or the lysosome (lumen) with a certainty of 0.1000.

Table 7D. Encoded NOV7b protein sequence (SEQ ID NO:38).

MELQDLELQLEERLLGLEEQLRVRMPSPFRSSALMGMCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSE LGLGLGEMGFEPGESSESVFSQATSESSSVCSGSPSHANRRRTGVPSTVSVGKSKTPLVARKKVFRASVALT PTAPSRTGSVQTPPDLESSEEVDAEAGAPEVVGPKSEVEEGHGKLPMPAVEEMHKNVQDELQQVIREIKE SIVGEIRREIVSGLLAUVSSSKASNSKQDYH

A search of sequence databases reveals that the NOV7b amino acid sequence has 240 of 249 amino acid residues (96%) identical to, and 242 of 249 amino acid residues (97%) similar to, the 249 amino acid residue ptnr:SWISSNEW-ACC:P28290 protein from *Homo sapiens* (Human) (Sperm-Specific Antigen 2 (Cleavage Signal-1 Protein) (CS-1)) ($E = 9.7e^{-121}$).

NOV7b is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Aorta, Ascending Colon, Bone, Cervix, Cochlea, Colon, Dermis, Gall Bladder, Hypothalamus, Islets of Langerhans, Liver, Lung, Lymphoid tissue, Ovary, Parathyroid Gland, Parotid Salivary glands, Pineal Gland, Retina, Right Cerebellum, Skin, Tonsils, Umbilical Vein, Vein, Whole Organism. .

NOV7c

A disclosed NOV7c nucleic acid of 806 nucleotides (also referred to as CG56613-03) encoding a novel cleavage signal-1 protein-like protein is shown in Table 7E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 21-23 and ending with a TAA codon at nucleotides 762-764. A putative untranslated region upstream from the initiation codon is underlined in Table 7E. The start and stop codons are in bold letters, and the 5' and 3' untranslated regions, if any, are underlined.

Table 7E. NOV7c nucleotide sequence (SEQ ID NO:39).

GTTTGAGAAATTCAGTCCGAATGGAACCTCAGGACCTGGAACCTGCAGCTGGAGGAGCGCCTGCTGGGCCTGG
 AGGAGCAGCTTCGTGCTGTGCGCATGCCCTTCACCTTCCGCTCCTCCGCACTCATGGGAATGTGTGGCAGTA
 GAAGCGCTGATAAATTGTCATGCCCTTCTCCATTGAATGTAATGGAACCACTGCACTGAACTGATGCAGGAGC
 AGTCATACCTGAAGTCTGAATTGGGCCTGGGACTTGGAGAAATGGGATTTGAAATTCCTCCTGGAGAAAGCT
 CAGAATCTGTTTTTCCCAAGCAACATCAGAATCATCTTCTGTATGTTCTGGTCCCTCTCATGCTAACAGAA
 GAACTGGAGTACCTTCTACTGCCTCAGTGGGCAAATCCAAAACCCATTAGTGGCAAGGAAGAAAGTGTTC
 GAGCATCGGTGGCTCTAACGCCAACAGCTCCTTCTAGAACAGGCTCTGTGCAGACACCTCCAGATTGGA
 GTTCTGAGGAAGTTGATGCAGCTGAAGGAGCCCCAGAAGTTGTAGGACCTAAATCTGAAGTGGGAAGGGC
 ATGGAAAACCTCCCATCAATGCCAGCTGCTGAGGAAATGCATAAAATGTGGAGCAAGATGAGTTGCAGCAAG
 TCATACGGGAGATTAAAGAGTCTATTGTTGGGAAATCAGACGGGAAATTGTAAGTGGACTTTTGGCAGCAG
 TATCTTCAAGTAAAGCGTCTAATTCTAAGCAAGATTATCATTAAACAGAAATTATAGGTTGGCATGGATCCT
 ATTAGCTGTGTAAT

The disclosed NOV7c nucleic acid sequence, localized to chromosome 2, has 803 of 812 bases (98%) identical to a gb:GENBANK-ID:HUMCS1PA|acc:M61199.1 mRNA from *Homo sapiens* (Human cleavage signal 1 protein mRNA, complete cds) ($E = 1.2e^{-171}$).

5 The disclosed NOV7c polypeptide (SEQ ID NO:40) encoded by SEQ ID NO:39 has 247 amino acid residues and is presented in Table 7F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV7c has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500. Alternatively, NOV7f may also localize to the mitochondrial matrix space with a certainty of 0.1000, or the lysosome
 10 (lumen) with a certainty of 0.1000.

Table 7F. Encoded NOV7c protein sequence (SEQ ID NO:40).

MELQDLELQLEERLLGLEEQLRVRMPSPFRSSALMGMCGRSADNLSCPSPLNVMEPVTELMQEQSYLKSE
 LGLGLGEMGFEI PPGESSES VFSQATSESSSVCSGPSHANRRRTGVPSTASVGKSKTPLVARKKVFRASVALT
 PTAPSRTGVSQTPPDLESSEEVDAEAGAPEVVGPKSEVEEGHGKLPMPAAEEMHKNVEQDELQQVIREIKE
 SIVGEIRREIVSGLLA AVSSSKASNSKQDYH

15 A search of sequence databases reveals that the NOV7c amino acid sequence has 242 of 249 amino acid residues (97%) identical to, and 244 of 249 amino acid residues (97%) similar to, the 249 amino acid residue ptnr:SWISSNEW-ACC:P28290 protein from *Homo sapiens* (Human) (Sperm-Specific Antigen 2 (Cleavage Signal-1 Protein) (CS-1)) ($E = 1.4e^{-121}$).

20 NOV7c is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Aorta, Ascending Colon, Bone, Cervix, Cochlea, Colon, Dermis, Gall Bladder, Hypothalamus, Islets of

Langerhans, Liver, Lung, Lymphoid tissue, Ovary, Parathyroid Gland, Parotid Salivary glands, Pineal Gland, Retina, Right Cerebellum, Skin, Tonsils, Umbilical Vein, Vein, Whole Organism. .

NOV7d

A disclosed NOV7d nucleic acid of 705 nucleotides (also referred to as 174307820) encoding a novel cleavage signal-1 protein-like protein is shown in Table 7G. An open reading frame was identified beginning with an AGA initiation codon at nucleotides 1-3 and ending with nucleotides 703-705. The start codon is in bold letters, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon is not a traditional initiation codon, and there is no stop codon, NOV7d could be a partial open reading frame extending further in the 5' and 3' directions.

Table 7G. NOV7d nucleotide sequence (SEQ ID NO:41).

AGATCTCCACCATGGAACCTCAGGACCTGGAACCTGCAGCTGGAGGAGCGCTGCTGGGCCTGGAGGAGCAG
CTTCGTGCTGTGCGCATGCCTTCACCCTTCCGCTCCTCCGCACTCATGGGAATGTGTGGCAGTAGAAGCGCT
GATAACTTGTCTATGCCCTTCTCCATTGAATGTAATGGAACCACTCACTGAACTGATGCAGGAGCAGTCATAC
CTGAAGTCTGAATTGGGCCTGGGACTTGGAGAAATGGGATTTGAAATTCCTCTGGAGAAAGCTCAGAATCT
GTTTTTCCCAAGCAACATCAGAATCATCTTCTGTATGTTCTGGTCCCTCTCATGCTAACAGAAGAACTGGG
GTACCTTCTACTGCCTCAGTGGGCAAATCCAAAACCCATTAGTGGCAAGGAAGAAAGTGTTCGGAGCATCG
GTGGCTCTAACGCCAACAGCTCCTTCTAGAACAGGCTCTGTGCAGACACCTCCAGATTGGAAAGTTCTGAG
GAAGTTGATGCAGCTGAAGGAGCCCCAGAAGTTGTAGGACCTAAATCTGAAGTGAAGAAGGGCATGGAAAA
CTCCCATCAATGCCAGCTGCTGAGGAAATGCATAAAATGTGGAGCAAGATGAGTTGCAGCAAGTCATACGG
GAGATTAAAGAGTCTATTGTTGGGGAAATCAGACGGGAAATTGTAAGTGGACTCGAG

The disclosed NOV7d polypeptide (SEQ ID NO:42) encoded by SEQ ID NO:41 has 235 amino acid residues and is presented in Table 7H using the one-letter amino acid code.

Table 7H. Encoded NOV7d protein sequence (SEQ ID NO:42).

RSPTMELQDLELQLEERLLGLEEQLRVRMPSPFRSSALMGMCGRSADNLSCPSPLNVMEPVTELMQEQSY
LKSELGLGLGEMGFEIPGESSESVFQSQATSESSVCSGPSHANRRTGVPSTASVGKSKTPLVARKKVFRAS
VALTPAPSRTGVSQTPPDLESSEEVDAEAGAPEVVGPKSEVEEGHGKLPMPAAEEMHKNVEQDELQQVIR
EIKESIVGEIRREIVSGLE

NOV7e

A disclosed NOV7e nucleic acid of 759 nucleotides (also referred to as 174307820) encoding a novel cleavage signal-1 protein-like protein is shown in Table 7I. An open reading frame was identified beginning with an AGA initiation codon at nucleotides 1-3 and ending with nucleotides 757-759. The start codon is in bold letters, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon is not a traditional initiation codon, and there is no stop codon, NOV7e could be a partial open reading frame extending further in the 5' and 3' directions.

Table 7I. NOV7e nucleotide sequence (SEQ ID NO:323).

AGATCTCCACCATGGAACCTCAGGACCTGGAAGTGCAGCTGGAGGAGCGCCTGCTGGGCTGGAGGAGCAG
CTTCGTGCTGTGCGCATGCCTTCACCCCTCCGCTCCTCCGCACTCATGGGAATGTGTGGCAGTAGAAGCGCT
GATAACTTGTTCATGCCCTTCTCCATTGAATGTAATGGAACCACTGACTGAAGTGCAGGAGCAGTCATAC
CTGAAGTCTGAATTGGGCTGGGACTTGGAGAAATGGGATTTGAAATTCCTCTGGAGAAAGCTCAGAATCT
GTTTTTCCCAAGCAACATCAGAATCATCTTCTGTATGTTCTGGTCCCTCTCATGCTAACAGAAGAACTGGA
GTACCTTCTACTGCCTCAGTGGGCAAATCCAAAACCCATTAGTGGCAAGGAAGAAAGTGTCCGAGCATCG
GTGGCTCTAACGCCAACAGCTCCTTCTAGAACAGGCTCTGTGCAGACACCTCCAGATTTGGAAAGTTCTGAG
GAAGTTGATGCAGCTGAAGGAGCCCCAGAAAGTTGTAGGACCTAAATCTGAAGTGGAGAAGGGCATGGAAAA
CTCCCATCAATGCCAGCTGCTGAGGAAATGCATAAAATGTGGAGCAAGATGAGTTGCAGCAAGTCATACGG
GAGATTAAAGAGTCTATTGTGGGAAATCAGACGGGAAATTGTAAGTGGACTTTTGGCAGCAGTATCTTCA
AGTAAAGCGTCTAATTCTAAGCAAGATTATCATCTCGAG

The disclosed NOV7e polypeptide (SEQ ID NO:324) encoded by SEQ ID NO:323 has 253 amino acid residues and is presented in Table 7J using the one-letter amino acid code.

Table 7J. Encoded NOV7e protein sequence (SEQ ID NO:324).

RSPTMELQDLELQLEERLLGLEEQRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQESY
LKSELGLGLGEMGFETPPGESSESVFSQATSESSSVCSGSPSHANRRRTGVPSTASVGKSKTPLVARKKVFRAS
VALTPTAPSRTGSVQTPPDLESSEVDAAEGAPEVVGPKSEVEEGHGKLPMPAAEEMHKNVEQDELQQVIR
EIKESIVGEIRREIVSGLLAUVSSSKASNSKQDYHLE

5. NOV7a also has homology to the amino acid sequence shown in the BLASTP data listed in Table 7K.

Table 7K. BLAST results for NOV7a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15620913 dbj BAB67820.1 (AB067514)	KIAA1927 protein [Homo sapiens]	772	242/247 (97%)	244/247 (97%)	e-109
gi 16159686 ref XP_057458.1 (XM_057458)	sperm specific antigen 2 [Homo sapiens]	727	242/247 (97%)	244/247 (97%)	e-108
gi 15277922 gb AAH12947.1 AAH12947 (BC012947)	Unknown (protein for MGC:21202) [Homo sapiens]	267	242/247 (97%)	244/247 (97%)	e-102
gi 5803179 ref NP_06742.1 (NM_006751)	sperm specific antigen 2; KIAA1927 protein [Homo sapiens]	249	247/249 (99%)	247/249 (99%)	e-102
gi 18017599 ref NP_542125.1 (NM_080558)	sperm specific antigen 2 [Mus musculus]	264	197/248 (79%)	212/248 (85%)	9e-81

- 10 The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 7L.

Table 7L. Information for the ClustalW proteins

- 15 1) NOV7a (SEQ ID NO:36)
2) NOV7b (SEQ ID NO:38)
3) NOV7c (SEQ ID NO:40)
4) NOV7d (SEQ ID NO:42)
5) NOV7e (SEQ ID NO:324)
5) gi|15620913|dbj|BAB67820.1| (AB067514) KIAA1927 protein [Homo sapiens] (SEQ ID NO:347)

- 6) gi|16159686|ref|XP_057458.1| (XM_057458) sperm specific antigen 2 [*Homo sapiens*]
(SEQ ID NO:348)
7) gi|15277922|gb|AAH12947.1|AAH12947 (BC012947) Unknown (protein for MGC:21202)
[*Homo sapiens*] (SEQ ID NO:349)
5 8) gi|5803179|ref|NP_006742.1| (NM_006751) sperm specific antigen 2; KIAA1927
protein [*Homo sapiens*] (SEQ ID NO:350)
9) gi|18017599|ref|NP_542125.1| (NM_080558) sperm specific antigen 2 [*Mus musculus*]
(SEQ ID NO:351)

10

		10	20	30	40	50	60
	NOV7a	1	1
	NOV7b	1	-----	-----	-----	-----	1
15	NOV7c	1	-----	-----	-----	-----	1
	NOV7d	1	-----	-----	-----	-----	1
	NOV7e	1	-----	-----	-----	-----	1
	gi 15620913	1	YSHIFLFDHLLRTASQHS	DSSGFAEDSTDCLSLNHLQVQESLQAMGSSADSCDSETT	VT	60	
	gi 16159686	1	-----	-----	-----	MGSSADSCDSETT	15
20	gi 15277922	1	-----	-----	-----	-----	1
	gi 5803179	1	-----	-----	-----	-----	1
	gi 18017599	1	-----	-----	-----	-----	1

25

		70	80	90	100	110	120
	NOV7a	1	1
	NOV7b	1	-----	-----	-----	-----	1
	NOV7c	1	-----	-----	-----	-----	1
	NOV7d	1	-----	-----	-----	-----	1
30	NOV7e	1	-----	-----	-----	-----	1
	gi 15620913	61	SLGEDLATPTAQDQPYFNESEESLVPLQKGLEKAAAVADKRKSGSQDFPQCNTIENTGT	120			
	gi 16159686	16	SLGEDLATPTAQDQPYFNESEESLVPLQKGLEKAAAVADKRKSGSQDFPQCNTIENTGT	75			
	gi 15277922	1	-----	-----	-----	-----	1
	gi 5803179	1	-----	-----	-----	-----	1
35	gi 18017599	1	-----	-----	-----	-----	1

40

		130	140	150	160	170	180
	NOV7a	1	1
	NOV7b	1	-----	-----	-----	-----	1
	NOV7c	1	-----	-----	-----	-----	1
	NOV7d	1	-----	-----	-----	-----	1
	NOV7e	1	-----	-----	-----	-----	1
45	gi 15620913	121	KQSTCSPGDHIIIEITEVEEDLFPAETVELLREASAESDVGKSSSEFTQYTTTHILKSLA	180			
	gi 16159686	76	KQSTCSPGDHIIIEITEVEEDLFPAETVELLREASAESDVGKSSSEFTQYTTTHILKSLA	135			
	gi 15277922	1	-----	-----	-----	-----	1
	gi 5803179	1	-----	-----	-----	-----	1
	gi 18017599	1	-----	-----	-----	-----	1

50

		190	200	210	220	230	240
	NOV7a	1	1
	NOV7b	1	-----	-----	-----	-----	1
	NOV7c	1	-----	-----	-----	-----	1
55	NOV7d	1	-----	-----	-----	-----	1
	NOV7e	1	-----	-----	-----	-----	1
	gi 15620913	181	SIEAKCSDMSSSENTTGPPSSMDRVNTALQRAQMKVCSLSNQRMGRSLLKSKDLLKQRYLF	240			
	gi 16159686	136	SIEAKCSDMSSSENTTGPPSSMDRVNTALQRAQMKVCSLSNQRMGRSLLKSKDLLKQRYLF	195			
	gi 15277922	1	-----	-----	-----	-----	1
60	gi 5803179	1	-----	-----	-----	-----	1
	gi 18017599	1	-----	-----	-----	-----	1

65

		250	260	270	280	290	300
	NOV7a	1	1
	NOV7b	1	-----	-----	-----	-----	1
	NOV7c	1	-----	-----	-----	-----	1
	NOV7d	1	-----	-----	-----	-----	1
	NOV7e	1	-----	-----	-----	-----	1

5	gi 15620913	241	AKAGYPLRRSQSLPTTLLSPVRVSSSVNVR	300
	gi 16159686	196	AKAGYPLRRSQSLPTTLLSPVRVSSSVNVR	255
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
10	NOV7a	1	1
	NOV7b	1	-----	1
	NOV7c	1	-----	1
	NOV7d	1	-----	1
	NOV7e	1	-----	1
15	gi 15620913	301	MEHDGQSLVKSTIFISPSVKKEEAPQSEAPRVEECHHGRTP	360
	gi 16159686	256	MEHDGQSLVKSTIFISPSVKKEEAPQSEAPRVEECHHGRTP	315
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
20	NOV7a	1	1
	NOV7b	1	-----	1
	NOV7c	1	-----	1
	NOV7d	1	-----	1
	NOV7e	1	-----	1
25	gi 15620913	361	SIHSEWQERPLCEHTRTLSTHSVPNISGATCSAFASPF	420
	gi 16159686	316	SIHSEWQERPLCEHTRTLSTHSVPNISGATCSAFASPF	375
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
30	NOV7a	1	1
	NOV7b	1	-----	1
	NOV7c	1	-----	1
	NOV7d	1	-----	1
	NOV7e	1	-----	1
35	gi 15620913	421	AIEMQLRRVLHDIRNSLQNLQSYPMRGPDPAAAPYSTQKSSVLPLYENTFQELQVMRRS	480
	gi 16159686	376	AIEMQLRRVLHDIRNSLQNLQSYPMRGPDPAAAPYSTQKSSVLPLYENTFQELQVMRRS	435
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
40	NOV7a	1	1
	NOV7b	1	-----	1
	NOV7c	1	-----	1
	NOV7d	1	-----	1
	NOV7e	1	-----	1
45	gi 15620913	481	LNLFRQTQMMDLAMLRLQQTVMVYHMTTEERFEVDQLQGLRNSVR	540
	gi 16159686	436	LNLFRQTQMMDLAMLRLQQTVMVYHMTTEERFEVDQLQGLRNSVR	495
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
50	NOV7a	1	15
	NOV7b	1	-----	15
	NOV7c	1	-----	15
	NOV7d	1	-----	15
	NOV7e	1	-----	15
55	gi 15620913	541	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	600
	gi 16159686	496	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	555
	gi 15277922	1	-----	1
	gi 5803179	1	-----	1
	gi 18017599	1	-----	1
60	NOV7a	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7b	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7c	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7d	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7e	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
65	gi 15620913	541	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	600
	gi 16159686	496	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	555
	gi 15277922	36	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	95
	gi 5803179	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	gi 18017599	36	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	95
70	NOV7a	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7b	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7c	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7d	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75
	NOV7e	16	GLEEQQLRAVRMPSPFRSSALMGCGSRSDNLSCPSPLNVMEPVTELMQEQSYLKSELGL	75

			610	620	630	640	650	660	
5	NOV7a	76	GLGEMGF	IPPGE	SSSVFS	KQSESS	SCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
	NOV7b	76	GLGEMGF	IPPGE	SSSVFS	KQSESS	SCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
	NOV7c	76	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
	NOV7d	76	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
	NOV7e	76	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
10	gi 15620913	601	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 659
	gi 16159686	556	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 614
	gi 15277922	96	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 154
	gi 5803179	76	GLGEMGF	IPPGE	SSSVFS	KQSESS	SCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 134
	gi 18017599	96	GLGEMGF	IPPGE	SSSVFS	QATSESS	SVCSGP	SHANRR	T-GVPSTASVGKSKTPLVARK 151
15			670	680	690	700	710	720	
	NOV7a	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 192
	NOV7b	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 192
	NOV7c	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 192
20	NOV7d	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 192
	NOV7e	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 192
	gi 15620913	660	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 717
	gi 16159686	615	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 672
	gi 15277922	155	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 212
25	gi 5803179	135	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 194
	gi 18017599	152	KVFRASV	ALTPT	APSRTG	SVQTP	PDLESSE	EVDAAE	GAPEVVGPKSE--VEEGHGKLP 209
30			730	740	750	760	770		
	NOV7a	193	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 247
	NOV7b	193	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 247
	NOV7c	193	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 247
	NOV7d	193	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 247
	NOV7e	193	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 247
35	gi 15620913	718	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 772
	gi 16159686	673	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 727
	gi 15277922	213	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 267
	gi 5803179	195	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKASNSKQDYH 249
40	gi 18017599	210	PAAEEMH	KNVEQ	DELQQV	IREIKES	IVGEIR	REIVSGL	LAAVSSSKAPGPKQDSH 264

The cleavage signal-1 protein (CS-1), a doublet antigen comprised of approximately 14-kDa and 18-kDa proteins has been shown to be present on the surface of sperm of various mammalian species including humans. Polyclonal antibodies to CS-1 inhibit the early cleavage of fertilized eggs without apparently affecting sperm penetration and pronuclear formation. The human CS-1 cDNA has been cloned and expressed in vitro to obtain the recombinant protein (reCS-1) molecule. The CS-1 cDNA clone has been isolated by immunological screening of a human testis lambda gt11 cDNA library with mono-specific polyclonal antibody against CS-1. The cDNA is 1828 bp long; the start codon assigned to the first ATG (bp 98-100) encodes a protein with 249 amino acid residues terminating at TAA (bp 845-847).

XCS-1 is a maternally expressed gene product that is the *Xenopus* homologue of the human cleavage signal protein (CS-1). XCS-1 may play an important role in regulating mitosis during early embryogenesis in *Xenopus laevis*. XCS-1 transcripts have been detected in oocytes. During development the XCS-1 protein has been detected on the membrane and in

the nucleus of blastomeres. It has also been detected on the mitotic spindle in mitotic cells and on the centrosomes in interphase cells. Overexpression of myc-XCS-1 in *Xenopus* embryos results in abnormal mitoses with increased numbers of centrosomes, multipolar spindles, and abnormal distribution of chromosomes. Incomplete cytokinesis resulting in multiple nuclei
5 residing in the same cytoplasm with the daughter nuclei in different phases of the cell cycle has been observed. The phenotype depended on the presence of the N terminus of XCS-1 (aa 1-73) and a consensus NIMA kinase phosphorylation site (aa159-167). Mutations in this site affect the ability of the overexpressed XCS-1 protein to produce the phenotype.

The disclosed NOV7 nucleic acid of the invention encoding a Cleavage signal-1
10 protein-like protein includes the nucleic acid whose sequence is provided in Table 7A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 7A while still encoding a protein that maintains its Cleavage signal-1 protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose
15 sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or
20 derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV7 protein of the invention includes the Cleavage signal-1 protein-
25 like protein whose sequence is provided in Table 7B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 2 while still encoding a protein that maintains its Cleavage signal-1 protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 21 percent of the residues may be so changed.

30 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Cleavage signal-1 protein-like protein (NOV7) is a member of a "Cleavage signal-1 protein family". Therefore, the NOV7 nucleic acids and proteins identified here may be useful in potential therapeutic applications

implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV7 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in regulation of the cell cycle during early embryogenesis, and therefore may have potential application in the management of embryonic defects. Additionally, this antigen may also be involved in human immunoinfertility and therefore may have application in the treatment of infertility, and/or other diseases or pathologies.

NOV7 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV7 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV7 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV8

A disclosed NOV8 nucleic acid of 2838 nucleotides (also referred to as 153472451) encoding a novel Matriptase-like protein is shown in Table 8A. An open reading frame was identified beginning with an TAG initiation codon at nucleotides 8-10 and ending with a TGA codon at nucleotides 2279-2281. The start and stop codons are in bold letters in Table 8A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 8A. NOV8 nucleotide sequence (SEQ ID NO:43).

<p>GGGGACCATGGGGAGCGATCGGGCCCGCAAGGGCGGAGGGGGCCCGAAGGACTTCGGCGCGGGACTCAAGTA CAACTCCCGGCACGAGAAAGTGAATGGCTTGGAGGAAGGCGTGGAGTTCTGCCAGTCAACAACGTCAAGAA GGTGGAAAAGCATGGCCCCGGGCGCTGGGTGGTGCTGGCAGCCGTGCTGATCGGCCTCCTTGTGGTGGAGGA GGCCGAGCGCGTCATGGCCGAGGAGCGCGTAGTCATGCTGCCCCCGCGGCGCGCTCCCTGAAGTCCTTTGT GGTCACCTCAGTGGTGGCTTTCCCCACGGACTCCAAACAGTACAGAGGACCCAGGACAACAGCTGCAGCTT TGGCCTGCACGCCCCGCGTGTGGAGCTGATGCGCTTACCACGCCCGGCTTCCCTGACAGCCCCCTACCCCGC TCATGCCCGCTGCCAGTGGGCCCTGCGGGGGACGCCGACTCAGTGCTGAGCCTCACCTTCCGCAGCTTTGA CCTTGCGTCCTGCGACGAGCGCGGAGCGACCTGGTGACGGGTACAACACCCTGAGCCCCATGGAGCCCCA CGCCCTGGTGCAGTTGTGTGGCACCTACCCTCCCTCCTACAACCTGACCTTCACTCCTCCAGAACGTCTCT GCTCATCACACTGATAACCAACTGAGCGCGGCATCCCGCTTTGAGGCCACCTTCTTCCAGCTGCCTAG</p>
--

GATGAGCAGCTGTGGAGGCCGCTTACGTAAAGCCCAGGGGACATTCAACAGCCCCTACTACCCAGGCCACTA
 CCCACCCAACTTGACTGCACATGGAACATTGAGGTGCCCAACAACAGCATGTGAAGGTGAGCTTCAAATT
 CTTCTACCTGCTGGAGCCCGCGTGCCTGCGGGCACCTGCCCAAGGACTACGTGGAGATCAATGGGGAGAA
 ATACTGCGGAGAGAGGTCCAGTTTCGTCTCACCAGCAACAGCAACAAGATCACAGTTTCGCTTCCACTCAGA
 TCAGTCTTACACCGACACCGGCTTCTTAGCTGAATACCTCTCTACGACTCCAGTGACCCATGCCCGGGGCA
 GTTACGTGCGGCACGGGGCGGTGTATCCGGAAGGAGCTGCGCTGTGATGGCTGGGCCGACTGCACCGACCA
 CAGCGATGAGCTCAACTGCAGTTGCGACGCGCGCCACAGTTTACGTGCAAGAACAAGTTTCTGCAAGCCCT
 CTTCTGGGTCTGCGACAGTGTGAACGACTGCGGAGACAACAGCGACGAGCAGGGGTGCAGTTGTCCGCCCCA
 GACCTTCAGGTGTTCCAATGGGAAGTGCCTCTCGAAAAGCCAGCAGTGCAATGGGAAGGACGACTGTGGGGA
 CGGGTCCGACGAGGCCTCTGCCCAAGGTGAACGTCGTCACTTGTACCAACACACCTACCGCTGCCTCAA
 TGGGCTCTGCTTGAGCAAGGGCAACCCTGAGTGTGACGGAAGGAGGACTGTAGCGACGGCTCAGATGAGAA
 GGACTGCGACTGTGGGCTGCGGTCTTACGAGACAGGCTCGTGTGTTGGGGGCACGGATGCGGATGAGGG
 CGAGTGGCCCTGGCAGGTAAGCCTGCATGCTCTGGGCCAGGGCCACATCTGCGGTGCTTCCCTCATCTCTCC
 CAACTGGCTGGTCTCTGCCGACACCTGCTACATCGATGACAGAGGATTACAGGTACTCAGACCCACGAGTG
 GACGGCCTTCTGGGCTTGCACGACACAGAGCCAGCGCAGCGCCCTGGGGTGCAGGAGCGCAGGCTCAAGCG
 CATCATCTCCACCCCTTCTCAATGACTTACCTTCGACTATGACATCGCGCTGCTGGAGCTGGAGAAACC
 GGCAGAGTACAGCTCCATGGTGGCGCCCATCTGCTGCCGACGCTCCCATGTCTTCCCTGCCGCAAGGC
 CATCTGGGTACCGGCTGGGGACACACCCAGTATGGAGGCACTGGCGCGCTGATCTGCAAAAGGGTGAGAT
 CCGCGTCTATCAACCAGACCACCTGCGAGAACCTCTGCGCAGCAGATCACGCCGCGCATGATGTGCGTGGG
 CTTCTCAGCGGCGCGTGGACTCTGCGCAGGGTGATTCCGGGGGACCCCTGTCCAGCGTGGAGGCGGATGG
 GCGGATCTTCCAGGCCGGTGTGGTGAGCTGGGGAGACGGCTGCGCTCAGAGGAACAAGCCAGGCGTGTACAC
 AAGGCTCCCTCTGTTTCGGGACTGGATCAAAGAGAACACTGGGGTATAGGGGCCGGGGCCACCCAAATGTGT
 ACACCTGCGGGGCCACCCATCGTCCACCCAGTGTGCACGCTGCAGGCTGGAGACTGGACCGCTGACTGCA
 CCAGCGCCCCCAGAACATACACTGTGAACCTCACTCCAGGGCTCCAAATCTGCCTAGAAAACCTCTCGCTT
 CCTCAGCCTCCAAAGTGGAGCTGGGAGGTAGAAGGGAGGACACTGGTGGTTCTACTGACCCAACTGGGGC
 AAAGGTTTGAAGACACAGCCTCCCCGCCAGCCCCAAGCTGGGCCGAGGCGCTTTGTGTATATCTGCCTCC
 CCTGTCTGTAAGGAGCAGCGGGAACGAGCTTCGGAGCCTCCTCAGTGAAGGTGGTGGGGCTGCCGGATCTG
 GGCTGTGGGGCCCTTGGGGCACGCTCTTGAAGAACCCAGGCTCGGAGGACCTGGAAAACAGACGGGTCTG
 AGACTGAAATTGTTTACCAGCTCCAGGGTGGACTTCAGTGTGTGTATTTGTGTAAATGGGTAAACAATT
 TATTTCTTTTAAAAA

The disclosed NOV8 nucleic acid sequence has 2644 of 2678 bases (98%) identical to a gb:GENBANK-ID:AF118224|acc:AF118224.2 mRNA from *Homo sapiens* (matriptase mRNA, complete cds) (E = 0.0).

- 5 The disclosed NOV8 polypeptide (SEQ ID NO:44) encoded by SEQ ID NO:43 has 757 amino acid residues is presented in Table 8B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV8 has a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.8110. Alternatively, NOV8 is predicted to be localized to the Golgi body with a certainty of 0.3000, to the endoplasmic
- 10 reticulum (membrane) with a certainty of 0.2000, or to the microbody (peroxisome) with a certainty of 0.1527. The most likely ceavage site for NOV8 is between positions 8 and 9, ARK-GG.

Table 8B. Encoded NOV8 protein sequence (SEQ ID NO:44).

MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPRWVVLAAVLIGLLVVEEAE
 RVMAEERVVMLPPRARSLKSFVVTSVVAFPTDSKTQVQRTQDNSSCSFGLHARGVELMRFTTPGFPDSPYPAHA
 RCQWALRGDADSVLSLTFRFDLASCDEGSDLVTVYNTLSPEMPHALVQLCGTYPSPYNLTFHSSQNVLLI
 TLIITNERRHPGFTEATFFQLPRMSSCGGRLRKAQGTFNSPYPGHPYNNIDCTWNIIEVPNNQHVKVSFKFFY
 LLEPGVPAGTCKPDYVEINGEKYCGERSQFVVTNSNKTIVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFT
 CRTGR CIRKELRCDGWADCTDHSDELNCS CDAGHQFTCKNKFKCLFWVCD SVNDCGDN SDEQGCSCPAQTF
 RCSNGKCLSKSQCNKGDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEKDC
 DCLRSFTRQARVVGGTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYSPTQWTA
 FLGLHDQSQRSAPGVQERRLKRIISHPFNFDTFDYDIALLELEKPAEYSSMVRPICLPDASHVFPAGKAIW

VTGWGHTQYGGTGALILQKGEIRVINQTTTCENLLPQQITPRMCMVGFLSGGVDSCQGDSSGGPLSSVEADGRI
FQAGVVSWDGCAQRNKPGVYTRLPLFRDWIKENTGV

A BLASTX of NOV8 shows that it has 699 of 729 amino acid residues (95%) identical to, and 702 of 729 amino acid residues (96%) similar to, the 855 amino acid residue ptnr:SPTREMBL-ACC:Q9Y5Y6 protein from *Homo sapiens* (Human) (Matriptase) (E = 0.0).

- 5 NOV8 is predicted to be expressed in at least the following tissues: Adrenal Gland/Suprarenal gland, Aorta, Ascending Colon, Bone Marrow, Brain, Bronchus, Cartilage, Colon, Duodenum, Gall Bladder, Heart, Islets of Langerhans, Kidney, Kidney Cortex, Lung, Mammary gland/Breast, Ovary, Pancreas, Parathyroid Gland, Parotid Salivary glands, Peripheral Blood, Pituitary Gland, Placenta, Prostate, Small Intestine, Stomach,
- 10 Thymus,Thyroid, Tonsils, Uterus, Vulva, Whole Organism.

In addition, NOV8 is predicted to be expressed in breast cancer, according to NOV8 nucleic acids, polypeptides, and antibodies. Accordingly to the invention will have diagnostic and therapeutic applications for the detection of breast cancer.

- 15 The disclosed NOV8 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 8C.

Table 8C. BLAST results for NOV8					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 10257390 gb AAG1 5395.1 AF057145_1 (AF057145)	serine protease TADG15 [<i>Homo sapiens</i>]	855	691/691 (100%)	691/691 (100%)	0.0
gi 11415040 ref NP_ 068813.1 (NM_021978)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [<i>Homo sapiens</i>]	855	690/691 (99%)	690/691 (99%)	0.0
gi 12249015 dbj BAB 20376.1 (AB030036)	prostamin [<i>Homo sapiens</i>]	855	689/691 (99%)	689/691 (99%)	0.0
gi 7363445 ref NP_0 35306.2 (NM_011176)	protease, serine, 14 (epithin) [<i>Mus musculus</i>]	855	573/691 (82%)	633/691 (90%)	0.0
gi 16758444 ref NP_ 446087.1 (NM_053635)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [<i>Rattus norvegicus</i>]	855	571/691 (82%)	632/691 (90%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 8D. In the ClustalW alignment of the NOV8 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 8D. ClustalW Analysis of NOV8

10	1) Novel NOV8 (SEQ ID NO:44)
	2) gi 10257390 gb AAG15395.1 AF057145_1 (AF057145) serine protease TADG15 [<i>Homo sapiens</i>] (SEQ ID NO:352)
	3) gi 11415040 ref NP_068813.1 (NM_021978) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [<i>Homo sapiens</i>] (SEQ ID NO:353)
15	4) gi 12249015 dbj BAB20376.1 (AB030036) prostamin [<i>Homo sapiens</i>] (SEQ ID NO:354)
	5) gi 7363445 ref NP_035306.2 (NM_011176) protease, serine, 14 (epithin) [<i>Mus musculus</i>] (SEQ ID NO:355)
20	6) gi 16758444 ref NP_446087.1 (NM_053635) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [<i>Rattus norvegicus</i>] (SEQ ID NO:356)
<div> <div>1020</div> <div>10</div> <div>20</div> <div>30</div> <div>40</div> <div>50</div> <div>60</div> </div>	
	NOV8 1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
25	gi 10257390 1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
	gi 11415040 1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
	gi 12249015 1 MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVVLAA 60
	gi 7363445 1 MGSNRGRKAGGGSQDFGAGLKYNSRLNMGFEEGVEFLPVNNAKKVEKRGPRRWVVLVA 60
30	gi 16758444 1 MCNNRGRKAGGGSQDFGAGLKYNSRLNMGFEEGVEFLPVNNAKKVEKRGPRRWVVLVA 60
<div> <div>70</div> <div>80</div> <div>90</div> <div>100</div> <div>110</div> <div>120</div> </div>	
	NOV8 61 VLIGLL 66
35	gi 10257390 61 VLIGLLVLLGIGFLVWHLQYRDVRVQKVFNCGYMRITNENFVDAYENSNSTEFVSLASKV 120
	gi 11415040 61 VLIGLLVLLGIGFLVWHLQYRDVRVQKVFNCGYMRITNENFVDAYENSNSTEFVSLASKV 120
	gi 12249015 61 VLIGLLVLLGIGFLVWHLQYRDVRVQKVFNCGYMRITNENFVDAYENSNSTEFVSLASKV 120
	gi 7363445 61 VLFSFLLLSLMAGLLVWFHRYRNVRIQKVFNCHLRITNEIFDAYENSTSTEFVSLASKV 120
40	gi 16758444 61 VLFSFLLLSLMAGLLVWFHRYRNVRIQKVFNCHLRITNENFVDAYENSTSTEFVSLASKV 120
<div> <div>130</div> <div>140</div> <div>150</div> <div>160</div> <div>170</div> <div>180</div> </div>	
	NOV8 66 LVEEAERVMAEERVVM 82
45	gi 10257390 121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
	gi 11415040 121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
	gi 12249015 121 KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERVMAEERVVM 180
	gi 7363445 121 KEALKLLYNEVPVLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEVDRAVMAERVVT 180
50	gi 16758444 121 KEALKLLMYSEVPVLGPYHKESAVTAFSEGSVIAYYWSEFSIPQHLVEEVDRAVMAERVVT 180
<div> <div>190</div> <div>200</div> <div>210</div> <div>220</div> <div>230</div> <div>240</div> </div>	
	NOV8 83 LPPRARSLKSFVVTSSVAFPTDSKTQVQRTQDNNSCSFGLHARGVELMRFTTPGFPDSPYPA 142
55	gi 10257390 181 LPPRARSLKSFVVTSSVAFPTDSKTQVQRTQDNNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
	gi 11415040 181 LPPRARSLKSFVVTSSVAFPTDSKTQVQRTQDNNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
	gi 12249015 181 LPPRARSLKSFVVTSSVAFPTDSKTQVQRTQDNNSCSFGLHARGVELMRFTTPGFPDSPYPA 240
	gi 7363445 181 LPPRARALKSFVVTSSVAFPTDPRMLQRTQDNNSCSFALHARGCAAVTRFTTPGFPDSPYPA 240
55	gi 16758444 181 LPPRARALKSFVVTSSVAFPTDPRMLQRTQDNNSCSFALHARGRTVTRFTTPGFPDSPYPA 240
<div> <div>250</div> <div>260</div> <div>270</div> <div>280</div> <div>290</div> <div>300</div> </div>	

NOV8 143 HARCQWALRGDADSVLSLTFRSDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 202
 gi|10257390| 241 HARCQWALRGDADSVLSLTFRSDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
 gi|11415040| 241 HARCQWALRGDADSVLSLTFRSDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
 gi|12249015| 241 HARCQWALRGDADSVLSLTFRSDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPPS 300
 5 gi|7363445| 241 HARCQWVLRGDADSVLSLTFRSDVAPCDHGSDLVTVYDLSLSPMEPHAVVRLCGTFSPS 300
 gi|16758444| 241 HARCQWVLRGDADSVLSLTFRSDVAPCDGHDSDLVTVYDLSLSPMEPHAVVRLCGTFSPS 300

310 320 330 340 350 360
 NOV8 203 YNLTFHSSQNVLLITLITNTERRHPGFEEATFFQLPRMSSCGRLRKAQGTFNSPYPYPGHY 262
 gi|10257390| 301 YNLTFHSSQNVLLITLITNTERRHPGFEEATFFQLPRMSSCGRLRKAQGTFNSPYPYPGHY 360
 gi|11415040| 301 YNLTFHSSQNVLLITLITNTERRHPGFEEATFFQLPRMSSCGRLRKAQGTFNSPYPYPGHY 360
 gi|12249015| 301 YNLTFHSSQNVLLITLITNTERRHPGFEEATFFQLPRMSSCGRLRKAQGTFNSPYPYPGHY 360
 10 gi|7363445| 301 YNLTFHSSQNVLLITLITNTDRHPGFEEATFFQLPRMSSCGFLSDTGTFFSPYPYPGHY 360
 15 gi|16758444| 301 YNLTFHSSQNVLLITLITNTDRHPGFEEATFFQLPRMSSCGFLSDTGTFFSPYPYPGHY 360

370 380 390 400 410 420
 NOV8 263 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 322
 gi|10257390| 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
 gi|11415040| 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
 gi|12249015| 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
 gi|7363445| 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420
 25 gi|16758444| 361 PPNIDCTWNIIEVPNNQHVKVSFKFFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVTS 420

430 440 450 460 470 480
 NOV8 323 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 382
 gi|10257390| 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480
 gi|11415040| 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480
 gi|12249015| 421 NSNKITVRFHSDQSYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480
 gi|7363445| 421 NSSKITVHFHSDHSDYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480
 30 gi|16758444| 421 NSSKITVHFHSDHSDYTDGTGFLAEYLSYDSSDPCPGQFTCRTRGRCIRKELRCDGWADCTDH 480

490 500 510 520 530 540
 NOV8 383 SDELNCSCDAGHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 442
 gi|10257390| 481 SDELNCSCDAGHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 540
 gi|11415040| 481 SDELNCSCDAGHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 540
 gi|12249015| 481 SDELNCSCDAGHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 540
 gi|7363445| 481 SDERHCRCNATHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 540
 40 gi|16758444| 481 SDERHCRCNATHQFTCKNKFKPLFWVCDVNDGDNSEQGCSCPAQTFRCNSNGKCLSK 540

550 560 570 580 590 600
 NOV8 443 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 502
 gi|10257390| 541 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600
 gi|11415040| 541 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600
 gi|12249015| 541 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600
 gi|7363445| 541 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600
 50 gi|16758444| 541 SQQCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600

610 620 630 640 650 660
 NOV8 503 DCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 562
 gi|10257390| 601 DCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 gi|11415040| 601 DCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 gi|12249015| 601 DCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 gi|7363445| 601 NCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 60 gi|16758444| 601 NCDCGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

670 680 690 700 710 720
 NOV8 563 DRGFRYSDPQTQWTAFLGLHDQSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 622
 gi|10257390| 661 DRGFRYSDPQTQWTAFLGLHDQSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
 gi|11415040| 661 DRGFRYSDPQTQWTAFLGLHDQSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
 gi|12249015| 661 DRGFRYSDPQTQWTAFLGLHDQSORSAPGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
 gi|7363445| 661 DKNFKYSDYTMWTAFLGLHDQSORSASGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720
 70 gi|16758444| 661 ETIFKYSDDHTMWTAFGLGLHDQSORSASGVQERRLKRIISHPFNDFTFDYDIALLELEKP 720

			730	740	750	760	770	780	
	NOV8	623	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE	682					
5	gi 10257390	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE	780					
	gi 11415040	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE	780					
	gi 12249015	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINOTTCE	780					
	gi 7363445	721	VEYSTVVRPICLPDASHVFPAGKAIWVTGWGHTKEGGTGALILQKGEIRVINOTTCE	780					
	gi 16758444	721	AEYSTVVRPICLPDASHVFPAGKAIWVTGWGHTKEGGTGALILQKGEIRVINOTTCE	780					
10			790	800	810	820	830	840	
	NOV8	683	PQQITPRMMCVGFLSGGVDS	742					
	gi 10257390	781	PQQITPRMMCVGFLSGGVDS	840					
	gi 11415040	781	PQQITPRMMCVGFLSGGVDS	840					
15	gi 12249015	781	PQQITPRMMCVGFLSGGVDS	840					
	gi 7363445	781	PQQITPRMMCVGFLSGGVDS	840					
	gi 16758444	781	PQQITPRMMCVGFLSGGVDS	840					
20			850						
	NOV8	743	RLPLFRDWIKENTGV	757					
	gi 10257390	841	RLPLFRDWIKENTGV	855					
	gi 11415040	841	RLPLFRDWIKENTGV	855					
	gi 12249015	841	RLPLFRDWIKENTGV	855					
25	gi 7363445	841	RLPVVRDWIKENTGV	855					
	gi 16758444	841	RLPEVRDWIKENTGV	855					

Tables 8E-8R list the domain descriptions from DOMAIN analysis results against NOV8. This indicates that the NOV8 sequence has properties similar to those of other proteins known to contain this domain.

Table 8E. Domain Analysis of NOV8

gnl|Smart|smart00020, Tryp_SpC, Trypsin-like serine protease; Many of these are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. (SEQ ID NO:804)
 CD-Length = 230 residues, 100.0% aligned
 Score = 259 bits (662), Expect = 4e-70

35	NOV 8:	516	RVVGGTDADEGEWPNQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFRYS	575
	Sbjct:	1	RIVGGSEANIGSFPWQVSLQYRGGRHFCGGSLSIPRWVLTAAHC-----VYGSAPSSIR	54
	NOV 8:	576	AFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKPAEYSSMVRPICLP	635
40	Sbjct:	55	VRLGSHDLS--SGEETQTVKSVKIVHPNPNSTYDNDIALLLKSEPVTLS	112
	NOV 8:	636	DASHVFPAGKAIWVTGWGHTQY-GGTGALILQKGEIRVINOTTCE	692
	Sbjct:	113	SSGYNVPAGTTCTVSGWGRTSESSGSLPDTLQEVNVPVSNATCRRAYS	172
45	NOV 8:	693	VGFLSGGVDS	751
	Sbjct:	173	AGGLEGGKDACQGD	230

gnl|Pfam|pfam00089, trypsin, Trypsin. Proteins recognized include all proteins in families S1, S2A, S2B, S2C, and S5 in the classification of peptidases. Also included are proteins that are clearly members, but that lack peptidase activity, such as haptoglobin and protein Z (PRTZ*). (SEQ ID NO:805)

CD-Length = 217 residues, 100.0% aligned
Score = 201 bits (510), Expect = 2e-52

5
10
15

gnl|Pfam|pfam00431, CUB, CUB domain (SEQ ID NO:806)
CD-Length = 110 residues, 100.0% aligned
Score = 99.0 bits (245), Expect = 9e-22

20
25

gnl|Pfam|pfam00431, CUB, CUB domain (SEQ ID NO:806)
CD-Length = 110 residues, 90.9% aligned
Score = 62.4 bits (150), Expect = 9e-11

30

35

Table 8I. Domain Analysis of NOV8

gnl|Smart|smart00042, CUB, Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein.; This domain is found mostly among developmentally-regulated proteins. Spermadhesins contain only this domain. (SEQ ID NO:807)
CD-Length = 114 residues, 99.1% aligned
Score = 97.4 bits (241), Expect = 3e-21

5
NOV 8: 242 CGGRLRKAQGTFNSPYYPGHYPNIDCTWNIEVPNNQHVKSFKFFYLLEPGVPAGTCPK 301
||| | + || || || |++| | | |+++ | | | + |
Sbjct: 1 CGGTLTASSGTITSPNYPNSYPNNLNCVWTISAPPGYRIELKFTDFDLE---SSDNCTY 56
NOV 8: 302 DYVEI-NGE-----KYCG-ERSQFVVTSSNSKITVRFHSDQSYTDTGFLAEYLS 348
|||| + | ++|| | +++|+|| +|| | || | || | | +
Sbjct: 57 DYVEIYDGPSTSSPLLGRFCGSELPPPIISSSSNSMTVTFVSDSSVQKRGSARYSA 113

Table 8J. Domain Analysis of NOV8

gnl|Smart|smart00042, CUB, Domain first found in C1r, C1s, uEGF, and bone morphogenetic protein.; This domain is found mostly among developmentally-regulated proteins. Spermadhesins contain only this domain. (SEQ ID NO:807)
CD-Length = 114 residues, 89.5% aligned
Score = 58.5 bits (140), Expect = 1e-09

10
NOV 8: 129 RFTTPGFDPSPYPAHARCQWALRGDADSVLSLTFRSFDLASCDESGDLVTVYNTLSPME 188
|+| +|+| || + | | + + | | || | | | | | +|+ |
Sbjct: 11 TITSPNYPNS-YPNNLNCVWTISAPPGYRIELKFTDFDLESSDNCTYDYVEIYDGPSTSS 69
15
NOV 8: 189 PHALVQLCGTYPPSYNLTFHSSQNVLITLITNTERRHPGFPEATFF 234
| | + ||+ | || | + +| +++++ + | | | +
Sbjct: 70 PL-LGRFCGSELP--PPIISSSSNSMTVTFVSDSSVQKRGSARYS 112

Table 8K. Domain Analysis of NOV8

gnl|Smart|smart00192, LDLa, Low-density lipoprotein receptor domain class A; Cysteine-rich repeat in the low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. The N-terminal type A repeats in LDL receptor bind the lipoproteins. Other homologous domains occur in related receptors, including the very low-density lipoprotein receptor and the LDL receptor-related protein/alpha 2-macroglobulin receptor, and in proteins which are functionally unrelated, such as the C9 component of complement. Mutations in the LDL receptor gene cause familial hypercholesterolemia. (SEQ ID NO:808)
CD-Length = 38 residues, 94.7% aligned
Score = 58.5 bits (140), Expect = 1e-09

20
NOV 8: 427 CPAQTFRCNSGKCLSKSQCCNGKDDCGDGSDEASCP 462
|| | +| ||+|+ | | +| ||||| ||| +||
Sbjct: 2 CPPGEFQCKNGRCIPLSWVCDGVDDCGDGSDEENCP 37

Table 8L. Domain Analysis of NOV8

gnl|Smart|smart00192, LDLa, Low-density lipoprotein receptor domain class A; Cysteine-rich repeat in the low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. The N-terminal type A repeats in LDL receptor bind the lipoproteins. Other homologous domains occur in related receptors, including the very low-density lipoprotein receptor and the LDL receptor-related protein/alpha 2-macroglobulin receptor, and in proteins which are functionally unrelated, such as the C9 component of complement. Mutations in the LDL receptor gene cause familial hypercholesterolemia. (SEQ ID NO:808)
CD-Length = 38 residues, 92.1% aligned
Score = 52.0 bits (123), Expect = 1e-07

NOV 8: 356 PGQFTCRTGRCIRKELRCDGWADCTDHSDELNCSC 390
||+| |+ ||| ||| || | ||| ||
Sbjct: 4 PGFQCKNGRCIPLSWVCDGVDDCGDGSDEENCPS 38

5

Table 8M. Domain Analysis of NOV8

gnl|Smart|smart00192, LDLa, Low-density lipoprotein receptor domain class A; Cysteine-rich repeat in the low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. The N-terminal type A repeats in LDL receptor bind the lipoproteins. Other homologous domains occur in related receptors, including the very low-density lipoprotein receptor and the LDL receptor-related protein/alpha 2-macroglobulin receptor, and in proteins which are functionally unrelated, such as the C9 component of complement. Mutations in the LDL receptor gene cause familial hypercholesterolemia. (SEQ ID NO:808)
CD-Length = 38 residues, 89.5% aligned
Score = 52.0 bits (123), Expect = 1e-07

NOV 8: 394 HQFTCKNKFKPLFWVCDVNDCCDNSDEQGCSC 427
+| ||| | || ||| |+||| |||+ |
Sbjct: 5 GEFQCKNGRCIPLSWVCDGVDDCGDGSDEENCPS 38

10

Table 8N. Domain Analysis of NOV8

gnl|Smart|smart00192, LDLa, Low-density lipoprotein receptor domain class A; Cysteine-rich repeat in the low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism. The N-terminal type A repeats in LDL receptor bind the lipoproteins. Other homologous domains occur in related receptors, including the very low-density lipoprotein receptor and the LDL receptor-related protein/alpha 2-macroglobulin receptor, and in proteins which are functionally unrelated, such as the C9 component of complement. Mutations in the LDL receptor gene cause familial hypercholesterolemia. (SEQ ID NO:808)
CD-Length = 38 residues, 94.7% aligned
Score = 45.1 bits (105), Expect = 1e-05

NOV 8: 468 TCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEKDC 504
|| ++| || |+ ||| +|| ||||++|
Sbjct: 1 TCPPGEFQCKNGRCIPLSWV-CDGVDDCGDGSDEENC 36

15

Table 8O. Domain Analysis of NOV8

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor
domain class A (SEQ ID NO:809)
CD-Length = 39 residues, 92.3% aligned
Score = 53.1 bits (126), Expect = 5e-08

NOV 8: 427 CPAQTFRCNSNGKCLSKSQQCNGKDDCGDGSDEASCP 462
| +| +|+| +| || |||| +|
Sbjct: 3 CGPNEFQCGSGECIPMSWVCDGDPDCEDGSDEKNCA 38

5

Table 8P. Domain Analysis of NOV8

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor
domain class A (SEQ ID NO:809)
CD-Length = 39 residues, 87.2% aligned
Score = 47.4 bits (111), Expect = 3e-06

NOV 8: 356 PGQFTCRTGRCIRKELRCDGWADCTDHSDELNCS 389
| +| +| || || || || || || +|
Sbjct: 5 PNEFQCGSGECIPMSWVCDGDPDCEDGSDEKNCA 38

10

Table 8Q. Domain Analysis of NOV8

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor
domain class A (SEQ ID NO:809)
CD-Length = 39 residues, 84.6% aligned
Score = 44.3 bits (103), Expect = 3e-05

NOV 8: 394 HQFTCKNKFKCKPLFWVCDSVNDCGDNSEQGC 426
++| | + | | + ||| || | ||| +| +|
Sbjct: 6 NEFQCGSGECIPMSWVCDGDPDCEDGSDEKNCA 38

15

Table 8R. Domain Analysis of NOV8

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor
domain class A (SEQ ID NO:809)
CD-Length = 39 residues, 92.3% aligned
Score = 42.0 bits (97), Expect = 1e-04

NOV 8: 468 TCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDGDEKDC 504
|| + ++| +| | + + ||| || ||||| +|
Sbjct: 2 TCGPNEFQCGSGECIPM-SWVCDGDPDCEDGSDEKNCA 37

20

The predicted sequence described here belongs to the leucine-rich repeat protein family. It is homologous to insulin like growth factor binding protein (IGFBP) and RP105, a novel B cell surface molecule. It contains five leucine-rich repeat domains. Leucine-rich repeats (LRRs) are relatively short motifs (22-28 residues in length) found in a variety of cytoplasmic, membrane and extracellular proteins (1). A common property of this protein family involves protein-protein interaction. Other functions of LRR-containing proteins

25

include, for example, binding to enzymes and vascular repair (1) . LRRs form elongated non-globular structures and are often flanked by cysteine rich domains. The circulating insulin-like growth factors (IGF-I and -II) occur largely as components of a 140kDa protein complex with IGF binding protein-3 and the acid-labile subunit (ALS). This ternary complex regulates the metabolic effects of the serum IGFs by limiting their access to tissue fluids.

Because of the presence of the Leucine rich repeat domains and the homology to the IGFBP and RP105, we anticipate that the novel sequence described here will have useful properties and functions similar to these genes.

The NOV8 nucleic acid and polypeptide contain structural motifs (i.e. leucine rich repeat domains) that are characteristics of proteins belonging to the leucine-rich repeat protein family. Accordingly, the various NOV8 nucleic acids and polypeptides of the invention are useful, inter alia, as novel members of this protein family.

The disclosed NOV8 nucleic acid of the invention encoding a Insulin like growth factor binding protein-like protein includes the nucleic acid whose sequence is provided in Table 8A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 8A while still encoding a protein that maintains its Insulin like growth factor binding protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acid, up to about 2 percent of the bases may be so changed.

The disclosed NOV8 protein of the invention includes the Insulin like growth factor binding protein-like protein whose sequence is provided in Table 8B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 8B while still encoding a protein that maintains its Insulin like growth factor binding protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Insulin like growth factor binding protein-like protein (NOV8) is a member of a "Insulin like growth factor binding protein family". Therefore, the NOV8 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV8 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diabetes, obesity, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, cirrhosis, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphoedema, and other diseases, disorders and conditions of the like.

NOV8 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV8 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV8 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV9

NOV9 includes three novel Neuropeptide Y/Peptide YY receptor -like proteins disclosed below. The disclosed sequences have been named NOV9a, and NOV9b.

NOV9a

A disclosed NOV9a nucleic acid of 2276 nucleotides (also referred to as CG56554-01) encoding a novel Neuropeptide Y/Peptide YY receptor -like protein is shown in Table 9A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 370-372 and ending with a TAA codon at nucleotides 1549-1551. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 9A. The start and stop codons are in bold letters.

Table 9A. NOV9a nucleotide sequence (SEQ ID NO:45).

GGCCAGAAGGCGGGGAGCCAGAGGCGGCAGGACCTAGCGTGGCGCTCCAGCACCCAGACCGTGGCGGGCGC
CTCGCCCTTAGGGAAGAGCAAGGGAAGAACTTTATTGTAACCGCGAACATTTTGGTCACTGAGATCGAGTC
TCCCAGTGCTTTGGCTTCCCGCTCTTTATCGTGGGTTTGATCCCTGAGCTGCTCTCCTTTCCCGAACCTCC
CGGGGTGCAGCCTAGAGCCCTCCCGCGCGGCTGACTCCAGAGTAGAGGAAGGGAGGCGGCTCCGGCTGGTC
CCCCGAAGCCCTCGCTGCCCCGAGATGCGGATGGCCAGCCAGTAGCGGGCGGTGGCCCCGCGTCCCGGGAG
CCACAGCAATGCAGGCGCTTAACATTACCCCGGAGCAGTTCTCTCGGCTGCTGCGGGACCACAACTGAGC
CGGGAGCAGTTCATCGCTCTGTACCGCTGCGACCGCTCGTCTACACCCAGAGCTGCGGGAGCGCCCAAG
CTGGCCCTCGTGCTCACCGGCGTGCTCATCTTCGCCCTGGCGCTCTTTGGCAATGCTCTGGTGTCTACGTG
GTGACCCGAGCAAGGCCATGCGCACCGTCACCAACATCTTATCTGCTCCTTGGCGCTCAGTGACCTGCTC
ATCACCTTCTTCTGCATTCCCGTCACCATGATCCAGAACATTTCCGACAACCTGGCTGGAGGGTGCTTTCATT
TGCAAGATGGTGCCATTTGTCCAGTCTACCGCTGTTGTGACAGAAATCCTCACTATGACCTGCATTGCTGTG
GAAAGGCACCAGGACTTGTGCATCCTTTTAAATGAAGTGGCAATACACCAACCGAAGGGCTTTCACAATG
CTAGGTGTGGTCTGGCTGGTGGCAGTCATCGTAGGATCACCCATGTGGCACGTGCAACAACTTGAGATCAAA
TATGACTTCCTATATGAAAAGGAACACATCTGCTGCTTAGAAGAGTGGACCAGCCCTGTGCACCAGAAGATC
TACACCACCTTCATCCTTGTCTATCCTCTTCCTCCTGCTCTTATGGAGAAGAAACGAGCTGTCTATTATG
GTGACAGTGGTGGCTCTCTTGTCTGTGTGCTGGGCACCATTCATGTTGTCCATATGATGATTGAATACAGT
AATTTTGAAAAGGAATATGATGATGTCAATCAAGATGATTTTGTCTATCGTGCAAAATATTGGATTTTTCC
AACTCCATCTGTAATCCCATGTCTATGCATTATGAATGAAAACCTTCAAAAAAATGTTTGTCTGCAGTT
TGTTATTGCATAGTAAATAAAACCTTCTCTCCAGCACAAAGGCATGGAAATTCAGGAATTACATGATGCGG
AAGAAAGCAAAGTTTTCCTCAGAGAGAATCCAGTGGAGGAAACCAAAGGAGAAGCATTCACTGATGGCAAC
ATTGAAGTCAAATGTGTGAACAGACAGAGGAGAAGAAAAGCTCAAACGACATCTTGCTCTCTTAGGTCT
GAACTGGCTGAGAATTCTCCTTTAGACAGTGGGCATTAAATTATAACAATATCTTCATAATTATGCCCTTCA
GATTGTAACCCAAAGAGAAAATTATTTGAGCAAAGGTCAAATACTCTTTTATTCTTAAGATGATGACAAG
AAGAAAACAAATCATGTTTCCATTAAAAATGACACGAGGCTAGTCCAAGTGCAGTGATGTTTACAACCAAT
TGATCACAAATCATTAAACAGATTCTGTGTTCTCTCATTCCCACTGCTTCACTTGACTAGCCTTAAAAAA
GCAACATGGAAGGCCAGGCACGGTGGCTCATGCTGTAATCCAGCACTTTGGGAGGCTAGACGGGCGGAT
CACGAGGTGAGGATCAAAACCATCTGGCTAACACGGTGAAACCCCATCTCTGCTAAAAATACAAAAATT
AGCCGGCGTGGTGGCGGGCACCTGTAGTCCAGCTACTTGGGAGCCTCAGGCGGGAGAATGGTGTGAACCC
GGGAGGCGGAGCTTGCAGTGATCCGAGATCATGCCACTGCACTCCAGCCTGGGCGAAGAGCGGAGTCCCC
GTCTCAAAAAAATTTTGTGAAAAATTCGTAACCATACTTTTAAGATTATTCAGTGGATTTTAAAAAT
CTGTACAGAAATCAGGGTTCTAGCTAGCAGTTTCTCCACGCACTGTAATGTGACTATGTATTG
CTAGATTGAATAAGAAAATAAAATAATATCTTCTCTCTGAAAA

In a search of public sequence databases, the NOV9a nucleic acid sequence, localized to chromosome 4, has 372 of 434 bases (85%) identical to a gb:GENBANK- ID:HSA400877|acc:AJ400877.1 mRNA from *Homo sapiens* (ASCL3 gene, CEGP1 gene, C11orf14 gene, C11orf15 gene, C11orf16 gene and C11orf17 gene) ($E = 2.5e^{-61}$).

The disclosed NOV9a polypeptide (SEQ ID NO:46) encoded by SEQ ID NO:45 has 393 amino acid residues and is presented in Table 9B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV9a has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV9a may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic

reticulum (membrane) with a certainty of 0.3000, or in the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV9a is between positions 64 and 65: GNA-LV.

Table 9B. Encoded NOV9a protein sequence (SEQ ID NO:46).

MQALNITPEQFSRLLRDHNLTREQFIALYRLRPLVYTPELPGRAKLALVLTGVLIFALALFGNALVFYVVTR SKAMRTVTNIFICSLALSDLLITFFCIPVTMIQNISDNWLEGAFICKMVPFVQSTAVVTEILTMTCIAVERH QGLVHPFKMKWQYTNRRFTMLGVVWLVAIVIGSPMWHVQOLEIKYDFLYEKEHICCLEEWTSPVHQKIYTT FILVILFLLPLMEKKRAVIMMVTVVALFAVCWAPFHVVHMMIEYSNFEKEYDDVTIKMIFAIVQIIGFSNSI CNPIVYAFMNENFKKNVLSAVCYCIVNKTFSAPQRHGNSGITMMRKKAKFSLRENPEETKGEAFSDGNIEV KLCEQTEKKKKLKRHLALFRSELAENSPLDSGH

A search of sequence databases reveals that the NOV9a amino acid sequence has 63 of 184 amino acid residues (34%) identical to, and 107 of 184 amino acid residues (58%) similar to, the 377 amino acid residue ptnr:SPTREMBL-ACC:O73733 protein from *Brachydanio rerio* (Zebrafish) (*Zebra danio*) (Neuropeptide Y/Peptide YY Receptor YA) (E = 0.0).

NOV9a is predicted to be expressed in at least kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in lower small intestine, colon, and pancreas, brain, hypothalamus because of SAGE tags identified for AI308124 and AI307658, ESTs which match to the sequence of the invention: pancreatic cancer, prostate, prostate cancer, brain, glioblastoma, astrocytoma, normal human luminal mammary epithelial cells, breast cancer, ovary, cystadenoma. The SAGE data is reproduced in Example 5. The sequence is also predicted to be expressed in the following tissues because of the expression pattern of related genes in the Neuropeptide Y/Peptide YY/ Orexin/ Galanin/ Cholecystokinin receptor family.

NOV9b

In the present invention, the target sequence identified previously, NOV9a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein

sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, 5 brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with 10 other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV9b. This differs from the previously 15 identified sequence (NOV9a) in having 38 less amino acids and 3 different ones.

A disclosed NOV9b nucleic acid of 1472 nucleotides (also referred to as CG56554-02) encoding a novel Neuropeptide Y/Peptide YY receptor -like protein is shown in Table 9C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 42- 44 and ending with a TAA codon at nucleotides 1335-1337. A putative untranslated region 20 upstream from the initiation codon and downstream from the termination codon is underlined in Table 9C. The start and stop codons are in bold letters.

Table 9C. NOV9b nucleotide sequence (SEQ ID NO:47).

CAGTAGCGGGCGGTGGCCCCGCGTCCCGGGAGCGCACAGCAATGCAGGCGCTTAACATTACCCCGGAGCAGT
TCTCTCGGCTGCTGCGGGACCACAACCTGACGCGGGAGCAGTTCAATCGCTCTGTACCGGCTGCGACCGCTCG
TCTACACCCAGAGCTGCGGGACGCGCAAGCTGGCCCTCGTGCTCACCGGCGTGCTCATCTTCGCCCTGG
CGCTCTTTGGCAATGCTCTGGTGTCTACGTGGTGACCCGCGAGCAAGGCCATGCGCACCGCTACCAACATCT
TTATCTGCTCCTTGGCGCTCAGTGACCTGCTCATCACCTTCTTCTGCATTCCCGTCACCATGCTCCAGAACA
TTTCCGACAACCTGGCTGGGGGGTGTCTTCATTGCAAGATGGTGCCATTGTGCCAGTCTACCGCTGTTGTGA
CAGAAATCCTCACTATGACCTGCATTGCTGTGGAAGGCCACGAGGACTTGTGCATCCTTTTAAATGAAGT
GGCAATACACCAACCGAAGGGCTTTACAATGCTAGGTGTGGTCTGGCTGGTGGCAGTCATCGTAGGATCAC
CCATGTGGCAGTGCAACAACCTTGAGATCAAATATGACTTCCTATATGAAAGGAACACATCTGCTGCTTAG
AAGAGTGGACAGCCCTGTGCACCAGAAGATCTACACCACCTTCATCCTTGTGCATCCTCTCTCCTGCTC
TTATGGTGATGCTTATTCTGTACAGTAAAATTGGTTATGAACCTTTGGATAAAGAAAAGAGTTGGGGATGGTT
CAGTGCTTCGAATATTATGGAAGAAATGTCCAAATAGCCAGGAAGAAGAAACGAGCTGTCATTATGA
TGGTGACAGTGGTGGCTCTCTTGTGTGTGTGGGCACCATTCATGTTGTCCATATGATGATGAATACA
GTAATTTTGAAGGAATATGATGATGTACAATCAAGATGATTTTGTCTATCGTGCAAAATTATGGATTTT
CCAACTCCATCTGTAATCCCATTTGTCTATGCATTTATGAATGAAACTTCAAAAAAATGTTTTGCTCGAG
TTTGTATATGCATAGTAAATAAACCTTCTCTCAGCACAAAGGCATGGAAATTGAGGAATTACAATGATGC
GGAAGAAAGCAAAGTTTTCCCTCAGAGAGAATCCAGTGAGGAAACCAAGGAGAAGCATTCAGTGATGGCA
ACATTGAAGTCAAATTGTGTGAACAGACAGAGGAGAAGAAAAGCTCAAACGACATCTTGCTCTCTTTAGGT
CTGAAGTGGCTGAGAATTCTCCTTTAGACAGTGGGCATTAAATTATAACAATATCTTCATAATTAAATGCCCTT
CAGATTGTAACCCAAAGAGAAAATTATTTGAGCAAAGGTCAAATACTCTTTTATTCTTAAGATGATGACA
AGAAGAAAACAAATATGTTTCATTAAAAATGA

In a search of public sequence databases, the NOV9b nucleic acid sequence, localized to chromosome 4, has 403 of 656 bases (61%) identical to a gb:GENBANK-ID:AB040103|acc:AB040103.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* OT7T022 mRNA for RFamide-related peptide receptor, complete cds) ($E = 7.8e^{-13}$).

The disclosed NOV9b polypeptide (SEQ ID NO:48) encoded by SEQ ID NO:47 has 393 amino acid residues and is presented in Table 9D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV9b has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV9b may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic reticulum (membrane) with a certainty of 0.3000, or in the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV9b is between positions 64 and 65: GNA-LV.

Table 9D. Encoded NOV9b protein sequence (SEQ ID NO:48).

MQALNITPEQFSRLLRDHNLTREQFIALYRLRPLVYTPELPGRAKLALVLTGVLI PALALFGNALVFYVTR SKAMRTVTNIFICSLALSDLLITFFCIPVTMIQNI SDNWLEGAFICKMVPFVQSTAVVTEILTMTCIAVERH QGLVHPFKMKWQYTNRRFTMLGVVWLVAIVGSPMWHVQ LEIKYDFLYEKEHICCLEEWTSPVHQKIYTT FILVILFLLPLMEKKRAVIMMVTVVALFAVCWAPFHV VHMMIEYSNFEKEYDDVTIKMIFAIVQIIIGFSNSI CNPIVYAFMNENFKKNVLSAVCYCIVNKTFSAPQRHNS GITMMRKKAKFSLRENPEETKGEAFSDGNIEV KLCEQTEKKKKLRHLALFRSELAENSPDLSGH

A search of sequence databases reveals that the NOV9b amino acid sequence has 108 of 315 amino acid residues (34%) identical to, and 180 of 315 amino acid residues (57%) similar to, the 522 amino acid residue ptrn:SWISSNEW-ACC:Q9Y5X5 protein from *Homo sapiens* (Human) (Neuropeptide Ff Receptor 2 (Neuropeptide G Protein-Coupled Receptor) (G-Protein-Coupled Receptor HLWAR77)) ($E = 5.2e^{-46}$).

NOV9b is predicted to be expressed in at least the following tissues: lower small intestine, colon, and pancreas, brain, hypothalamus, kidney, pancreatic cancer, prostate, prostate cancer, glioblastoma, astrocytoma, normal human luminal mammary epithelial cells, breast cancer, ovary, cystadenoma. .

The disclosed NOV9a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 9E.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16566347 gb AAL26488.1 AF411117_1 (AF411117)	G protein-coupled receptor [<i>Homo sapiens</i>]	455	382/393 (97%)	384/393 (97%)	0.0
gi 13027438 ref NP_076470.1 (NM_023980)	neuropeptide FF receptor 2 [<i>Rattus norvegicus</i>]	417	99/314 (31%)	157/314 (49%)	3e-37
gi 4106397 gb AAD02833.1 (AF073925)	neuropeptide Y/peptide YY receptor Yb [<i>Gadus morhua</i>]	374	90/320 (28%)	169/320 (52%)	4e-37
gi 4758820 ref NP_04876.1 (NM_004885)	neuropeptide G protein-coupled receptor; neuropeptide FF 2 [<i>Homo sapiens</i>]	522	98/317 (30%)	159/317 (49%)	4e-37
gi 13878604 sp Q9Y5X5 NFF2_HUMAN	NEUROPEPTIDE FF RECEPTOR 2 (NEUROPEPTIDE G PROTEIN-COUPLED RECEPTOR) (G-PROTEIN-COUPLED RECEPTOR HLWAR77)	522	98/317 (30%)	159/317 (49%)	4e-37

5

10

- 15

25

gi 13027438	1	-----	1
gi 4106397	1	-----	1
gi 4758820	1	MNSFFGTPAASWCLESVDVSSAPDKEAGRERRALSVQQRGGPAWSGSLEWSRQSAGDRRR	60
gi 13878604	1	MNSFFGTPAASWCLESVDVSSAPDKEAGRERRALSVQQRGGPAWSGSLEWSRQSAGDRRR	60

		70	80	90	100	110	120	
NOV9a	1	-----	-----	-----	-----	-----	-----	17
NOV9b	1	-----	-----	-----	-----	-----	-----	17
gi 16566347	20	TGIVLANSSLDIVLHDTYYVVAHCGGNVRLHCGGPASRERTAMQALNITPEQFSRLLRD	79					
gi 13027438	1	-----	-----	-----	-----	-----	-----	18
gi 4106397	1	-----	-----	-----	-----	-----	-----	11
gi 4758820	61	LGLSRQTAKSSWSRSRDRTCCCRRAWIILVPAADRARRERFIMNEKWDNTNSENWHP	120					
gi 13878604	61	LGLSRQTAKSSWSRSRDRTCCCRRAWIILVPAADRARRERFIMNEKWDNTNSENWHP	120					

		130	140	150	160	170	180	
NOV9a	18	HNLTRQFIALYRLRPLVYTPELEGRAKLALVLTGVLIFALALFENALVFYVVT	77					
NOV9b	18	HNLTRQFIALYRLRPLVYTPELEGRAKLALVLTGVLIFALALFENALVFYVVT	77					
gi 16566347	80	HNLTRQFIALYRLRPLVYTPELEGRAKLALVLTGVLIFALALFENALVFYVVT	139					
gi 13027438	19	GNDTQHPWYSIDINITYMNYLHQP-HVTAVFISSYFLIFFLCMVGNVTVCFV	77					
gi 4106397	12	SHPKAN--YSLIQAWDOEECPSSKSGTTFLLVYSTMIAGVIGNSCLVF	69					
gi 4758820	121	VNDTKHHLYSDINITYMNYLHQP-QVAAIFLISYFLIFFLCMVGNVTVCF	179					
gi 13878604	121	VNDTKHHLYSDINITYMNYLHQP-QVAAIFLISYFLIFFLCMVGNVTVCF	179					

		190	200	210	220	230	240	
NOV9a	78	TVTNIFICSIALSDLLITFFCIPVTMIQNISDNWLEGAFICKMVPFVQSTAVTE	137					
NOV9b	78	TVTNIFICSIALSDLLITFFCIPVTMIQNISDNWLEGAFICKMVPFVQSTAVTE	137					
gi 16566347	140	TVTNIFICSIALSDLLITFFCIPVTMIQNISDNWLEGAFICKMVPFVQSTAVTE	199					
gi 13027438	78	TVTNFFIFNLATSDLLVGFICMPTLLDNIHAGWPFSSMCKISGLVQGISVAASV	137					
gi 4106397	70	NVTNIFIANLSCSDILMCI FCLPVTLIYTLMDRWILGEALCKLTPFVQCISV	129					
gi 4758820	180	TVTNIFILNLATSDLLVGFICMPTLLDNIHAGWPFSSMCKISGLVQGISVAASV	239					
gi 13878604	180	TVTNIFILNLATSDLLVGFICMPTLLDNIHAGWPFSSMCKISGLVQGISVAASV	239					

		250	260	270	280	290	300	
NOV9a	138	CIAVERHOGLVHPFKMKWQYTNRRRAFTMLGVVWLVAIVGSP---MWHVQQL	194					
NOV9b	138	CIAVERHOGLVHPFKMKWQYTNRRRAFTMLGVVWLVAIVGSP---MWHVQQL	194					
gi 16566347	200	CIAVERHOGLVHPFKMKWQYTNRRRAFTMLGVVWLVAIVGSP---MWHVQQL	256					
gi 13027438	138	AIADVDRERCVMYVFPKP--KLTVKTAFTVMTIIVGLAITMTSPAIMLHVQ	195					
gi 4106397	130	LIAMERYQLITHPTGW--KPMVGQSYMAVGIIWVACLISVFLSFTVLIDNLP	187					
gi 4758820	240	AIADVDRFQCVVYVFPKP--KLTIKTAFTVMTIIVGLAITMTSPSAVMLHVQ	297					
gi 13878604	240	AIADVDRFQCVVYVFPKP--KLTIKTAFTVMTIIVGLAITMTSPSAVMLHVQ	297					

		310	320	330	340	350	360	
NOV9a	195	EKEHIC---CLEEWTSPVHQKIYTTFLVILFLPL	227					
NOV9b	195	EKEHIC---CLEEWTSPVHQKIYTTFLVILFLPL	227					
gi 16566347	257	EKEHIC---CLEEWTSPVHQKIYTTFLSSSSSCLL	289					
gi 13027438	196	SHNKTSTVYWCREDPNQEMRRIYTTVLFATYLAPLSLIVIMYARIGASLFT	255					
gi 4106397	188	FPGQDH--WLCTESWPTNSNRLAYTTSLLVFOFLPLGLIAACYLSIFLRL	244					
gi 4758820	298	SONKTSPVYWCREDPNQEMRRIYTTVLFATYLAPLSLIVIMYGRIGISLFR	357					
gi 13878604	298	SONKTSPVYWCREDPNQEMRRIYTTVLFATYLAPLSLIVIMYGRIGISLFR	357					

		370	380	390	400	410	420	
NOV9a	227	ME---KKRAVINMVTIVALEAVCWAPFHVVHMMIEYSNFEKEYDDVTIKM	276					
NOV9b	227	ME---KKRAVINMVTIVALEAVCWAPFHVVHMMIEYSNFEKEYDDVTIKM	276					
gi 16566347	289	WK---KKRAVINMVTIVALEAVCWAPFHVVHMMIEYSNFEKEYDDVTIKM	338					
gi 13027438	256	QRLEQWHVS--KKQKVIKMLLTVALLETLSWLPLWTIMMLSDYADLSPNEL	314					
gi 4106397	245	RARDSSRDNRAGSRRINVMGLSIVALEAVCWLPENIFNTVFDWH--HELMV	302					
gi 4758820	358	RKNQEQWHVVSRRKQKIKMLLIVALLELTLSWLPLWTIMMLSDYADLSPNEL	417					
gi 13878604	358	RKNQEQWHVVSRRKQKIKMLLIVALLELTLSWLPLWTIMMLSDYADLSPNEL	417					

		430	440	450	460	470	480	
NOV9a	277	AIVQITGFSNSICNPIVYAFMNFNFKKNVLSAVCYCTVKNKTFSPAORHGN	336					
NOV9b	277	AIVQITGFSNSICNPIVYAFMNFNFKKNVLSAVCYCTVKNKTFSPAORHGN	336					
gi 16566347	339	AIVQITGFSNSICNPIVYAFMNFNFKKNVLSAVCYCTVKNKTFSPAORHGN	398					

gi|13027438| 315 PFAHWLAFGNSSVNPITYGFFNENFRSGFQDAFQF--CQKKVKPQEAAYG-----LRKR 366
gi|4106397| 303 SVCHLVAMASTCVNPVVYGFNLNSNFQQLKATLSHC---RCWGAAERYEN----- 349
gi|4758820| 418 PFAHWLAFGNSSVNPITYGFFNENFRSGFQDAFQFQQLCQKRAKPMEAAYT-----LKAKS 471
gi|13878604| 418 PFAHWLAFGNSSVNPITYGFFNENFRSGFQDAFQFQQLCQKRAKPMEAAYA-----LKAKS 471

490 500 510 520 530
.....
NOV9a 337 KFSLRNPVEETKGEAFSDGNIEVKLCEQTEKKKLRHIALFRSELAENSPLDSGH 393
NOV9b 337 KFSLRNPVEETKGEAFSDGNIEVKLCEQTEKKKLRHIALFRSELAENSPLDSGH 393
10 gi|16566347| 399 KFSLRNPVEETKGEAFSDGNIEVKLCEQTEKKKLRHIALFRSELAENSPLDSGH 455
gi|13027438| 367 NLDIN-TSGLLVHEPASQNPSENLGCRKSADNPTQESLMEETGEATNSTET----- 417
gi|4106397| 349 -LPLSAVSTEVTKESHMSKGSISHS----- 374
gi|4758820| 472 HVLIN-TSNQIVQESTFONPHGETLLYRKSAEKPOQELVMEELKETTNSSEI----- 522
15 gi|13878604| 472 HVLIN-TSNQIVQESTFONPHGETLLYRKSAEKPOQELVMEELKETTNSSEI----- 522

Tables 9G-9H list the domain descriptions from DOMAIN analysis results against NOV9. This indicates that the NOV9 sequence has properties similar to those of other proteins known to contain this domain.

Table 9G Domain Analysis of NOV9

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 146 bits (368), Expect = 2e-36

NOV 9: 62 GNALVFYVVTRSKAMRTVTNIFICSLALSDLLITFFCIPVTMIQNISDNWLEGAFICKMV 121
25 Sbjct: 1 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPWALYYLVGGDWVFGDALCKLV 60
NOV 9: 122 PFVQSTAVVTEILTMTCIAVERHQGLVHPFKMKWQYTNRRRAFTMLGVVWLVAIVIGSPMW 181
30 Sbjct: 61 GALFVVNGYASILLTIAISIDRYLAIVHPLRYRIRTPRRAKVLILLVWVLALLLSLPPL 120
NOV 9: 182 HVQQLKIKYDFLYEKEHICCLEEWTSVPVHQKIYTTFILVILFLLPL----- 227
Sbjct: 121 LFSWLR---TVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPPLLVLVCYTRILRTL 176
35 NOV 9: 228 -----MEKKRAVIMMVTVVALFAVCWAPFHVVHMMIEYSNFEKEYDDVTIK 273
Sbjct: 177 RKRARSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLL---DSLCLLSIWRVLP 233
NOV 9: 274 MIFAIVQIIGFSNSICNPITY 294
40 Sbjct: 234 TALLITLWLAYVNSCLNPITY 254

Table 9H Domain Analysis of NOV9

gnl|Pfam|pfam01604, 7tm_5, 7TM chemoreceptor. This large family of proteins are related to pfam00001. They are 7 transmembrane receptors. This family does not include all known members, as there are problems with overlapping specificity with pfam00001. This family is greatly expanded in the nematode worm *C. elegans*. (SEQ ID NO:811)

CD-Length = 297 residues, 83.8% aligned

Score = 38.1 bits (87), Expect = 0.001

Table 9H Domain Analysis of NOV9

gnl|Pfam|pfam01604, 7tm_5, 7TM chemoreceptor. This large family of proteins are related to pfam00001. They are 7 transmembrane receptors. This family does not include all known members, as there are problems with overlapping specificity with pfam00001. This family is greatly expanded in the nematode worm *C. elegans*. (SEQ ID NO:811)

CD-Length = 297 residues, 83.8% aligned

Score = 38.1 bits (87), Expect = 0.001

	NOV 9:	55	IFALALFGNALVFYVVTR--SKAMRTVTN---IFICSLALSDLLITFFCIPVTMIQNISD	109
			++ + + ++ ++ ++	
5	Sbjct:	16	ITIISLPIHIFGFYCILFKTPKKMKSVKWSLLNLHFWSALLDLYLSFLTIPYLFPPVLAG	75
	NOV 9:	110	NWLEGAFIGCKMVPFVQSTAVVTEILTMTC---IAVERHQGLVHPFKMKWQYTNRRAFSTM	165
			+ + + + + + + ++	
	Sbjct:	76	YPLGLLSYLGVPSTSIQIYIGVTILGVVAVSIIILLFENRHSLSLVNINN-KFRIWKWIRILY	134
10	NOV 9:	166	LGVVVLVAVIVGSPMWHVQQLLEIKYDFLYEKEHICCLEEWTSPVHQKIYTTFILVILFLL	225
			+ +++ + ++ + + + ++ + + + + + + +	
	Sbjct:	135	LILNYILAVLFFLPVFLLPEDQEAALKLKKYPCPPPEFFDEPNFFVLAIIDSNYFVISI	194
	NOV 9:	226	PLMEKKRAVIMMVTVVALFAVCWAPFHVHMMIEYSNFEKEYDDVTIKMIFAI-VQIIGF	284
15			+ + + + + + + + + + + + + + + + + +	
	Sbjct:	195	VFLI---LIVILQIIFFVSLIFYYLKILKNSTMSKKTRKLQ-----KKFFIALCIQVSIP	246
	NOV 9:	285	SNSICNPIVYAFMNENFK 302	
			++ +	
20	Sbjct:	247	ILVILIPLIYLVFSIIFG 264	

The NOV9 nucleic acids and polypeptides share structure similarity to members to the Neuropeptide Y/Peptide YY/ Orexin/ Galanin/ Cholecystokinin/pancreatic polypeptide receptor family. Neuropeptide Y (NPY) is one of the most abundant neuropeptides in the mammalian nervous system and exhibits a diverse range of important physiologic activities, including effects on psychomotor activity, food intake, regulation of central endocrine secretion, and potent vasoactive effects on the cardiovascular system. It shows sequence homology to peptide YY and over 50% homology in amino acid and nucleotide sequence to pancreatic polypeptide. Neuropeptide Y (NPY) signals through a family of G protein-coupled receptors present in the brain and sympathetic neurons. At least 3 types of neuropeptide Y receptor have been defined on the basis of pharmacologic criteria, tissue distribution, and structure of the encoding gene. The NPY Y1 receptors have been identified in a variety of tissues, including brain, spleen, small intestine, kidney, testis, placenta, and aortic smooth muscle. The Y2 receptor is found mainly in the central nervous system.

35 Orexin A and Orexin B, are derived from the same precursor, orexin, or hypocretin (HCRT), by proteolytic processing. One receptor, designated OX2R, binds both orexin A and orexin B. The predicted amino acid sequences of human and rat OX2R are 95% identical and contain 7 putative transmembrane domains. The other receptor, designated OX1R (HCRT1), binds orexin A only and has 64% identity to OX2R. Northern blot analysis revealed that in the

rat a 3.5-kb OX2R mRNA is expressed exclusively in the brain. When administered intracerebroventricularly to rats, orexin A and orexin B stimulated food consumption. In addition, preproorexin mRNA levels are upregulated upon fasting, thus these peptides are mediators in the central feedback mechanism that regulates feeding behavior.

5 PYY is secreted from endocrine cells in the lower small intestine, colon, and pancreas. It acts through the pancreatic polypeptide receptors in the gastrointestinal tract as an inhibitor of gastric acid secretion, gastric emptying, digestive enzyme secretion by the pancreas, and gut motility.

The disclosed NOV9 nucleic acid of the invention encoding a Neuropeptide Y/Peptide YY receptor -like protein includes the nucleic acid whose sequence is provided in Table 9A or
10 a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 9A while still encoding a protein that maintains its Neuropeptide Y/Peptide YY receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
15 nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones
20 are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 15 percent of the bases may be so changed.

25 The disclosed NOV9 protein of the invention includes the Neuropeptide Y/Peptide YY receptor -like protein whose sequence is provided in Table 9B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 9B while still encoding a protein that maintains its Neuropeptide Y/Peptide YY receptor -like activities and physiological functions, or a functional fragment
30 thereof. In the mutant or variant protein, up to about 70 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Neuropeptide Y/Peptide YY receptor-like protein (NOV9) is a member of a "Neuropeptide Y/Peptide YY receptor family". Therefore, the NOV9 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV9 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in obesity, diabetes, kidney disorders, cardiovascular disorders, anorexia, eating disorders, gastrointestinal and digestive diseases, metabolic diseases, CNS disorders, cancer, autoimmune disease, inflammation, and/or other pathologies and disorders.

NOV9 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV9 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV9 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV10

A disclosed NOV10 nucleic acid of 985 nucleotides (also referred to as CG55964-01) encoding a novel G-Protein Coupled Receptor-like protein is shown in Table 10A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 33-35 and ending with a TGA codon at nucleotides 981-983. A putative untranslated region upstream from the initiation codon is underlined in Table 10A. The start and stop codons are in bold letters.

Table 10A. NOV10 nucleotide sequence (SEQ ID NO:49).

CAAAATCTACCACTTGATTCTGATGAACAAATCATGCGGACATTCAATGGCTCAGTCTTCATGCCCTCTGCGT
TTATACTAATTTGGGATTCTGGTCTGGAGTCAGTGCAGTGTGGATTGGGATTCTTTCTCTGCCATGTATC
TTATTGTGTGATTGGAAATCCCTAATTTTTAGTTATAAATCAAATATGAAACAGCCTCCATATACCCATG
ACATTTTTTTGGCCATGTGGCAGCACAGACATTTGCATCTTAACACCTGCATTCTTCCAAAATGTTAGGCA
TCTTCTGGTTTCATTTGCCAGAGATTTCTTTGATGCCTGTCTTTTCAAATGTGGCTTATTCACCTATTCC
AGGCAATTGAATCGGGTATCTCTTGGCAATGGCCCTGGATCGCTATGTGGCCATCTGTATCCCTTGGAGAC
ATGCCACCATCTTTTCCAGCAGTTCTTAACATAATTGGACTTGGGGTGACACTCAGGGCTGCCATTCTTA
TAATACCTTCCTTAGGGCTCATCAAATGCTGTCTGAAACACTATCGAACTACAGTCATCTCTCACTCTTACT
GTGAGCACATGGCCATCGTGAAGCTGGCTACTGAAGATATCCGAGTCAACAAGATATATGGCCCTATCTGGT
CCTTTGCAATCTAGGGTTTGACATAATATTTAAACCTTGCTATGTCCCAATTTTATCACTGTCTTTC
AGCTGCCCAAGCAAGGAGGCACGATTAAGGCCCTTTAATACATGTCATTGCCACATTTGTGTCTTCTACAGT
TCTACCTTCTTGCTTCTCTCTTTCTTTCACACAGGTTTGGTTTCACACATACCACATATATTCATATCC
TCTTGTCAAATCTTTACCTGTTAGTCTCCACCTTTCTCAACCTATTTGTCTATGGAGTGAAGACCAAGCAA
TCTCGTACCATTATTGTGAAGTGTGTTTTCTTCAAATAAGTAACCTTGATC

In a search of public sequence databases, the NOV10 nucleic acid sequence has 789 of 974 bases (81%) identical to a gb:GENBANK-ID:AF133300|acc:AF133300.2 mRNA from *Mus musculus* (MOR 3'Beta1, MOR 3'Beta2, MOR 3'Beta3, and MOR 3'Beta4 genes, complete cds; Cbx3 pseudogene, complete sequence; and MOR 3'Beta5 and MOR 3'Beta6 genes, complete cds) ($E = 4.3e^{-136}$).

The disclosed NOV10 polypeptide (SEQ ID NO:50) encoded by SEQ ID NO:49 has 316 amino acid residues and is presented in Table 10B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV10b has a signal peptide and is likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.6850. Alternatively, NOV10 may also localize to the plasma membrane with a certainty of 0.6400, the Golgi body with a certainty of 0.4600, or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV10 is between positions 24 and 25: LES-VO.

Table 10B. Encoded NOV10 protein sequence (SEQ ID NO:50).

MPTFNGSVFMPSAFILIGIPGLSEVQCWIGIPFSAMYLGIVIGNSLILVI IKYENSLHIPMYIF
 LAMLAATDIALNTCILPKMLGIFWFHLP EISFDACLFQMWLIIHSFQAIESGILLAMALDRYVAI
 CIPLRHATIPSQOFLTHIGLGVTLRAAILIIPSLGLFCKLKHRYRTTVISHSYCEHMAIVKLAT
 EDIRVNKIYGLFVAFAILGFDIIFITLSYVOIFITVQLPQKEARFAKENTCIAHICVFLQFYL
 LAFFSFTHRGSHIPPYIHILLSNLYLLVPPFLNP IVYGVKTKQIRDHIVKVVFFKKVT

A search of sequence databases reveals that the NOV10 amino acid sequence has 316 of 316 amino acid residues (100%) identical to, and 316 of 316 amino acid residues (100%) similar to, the 316 amino acid residue ptnr:TREMBLNEW-ACC:AAG42368 protein from *Homo sapiens* (Human) (Odorant Receptor HOR3'BETA5) ($E = 5.7e^{-169}$).

NOV10 is predicted to be expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV10 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 10C.

Table 10C. BLAST results for NOV10					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 11991867 gb AAG4 2368.1 (AF289204)	odorant receptor HOR3'beta5 [<i>Homo sapiens</i>]	316	316/316 (100%)	316/316 (100%)	e-148
gi 7305351 ref NP_0 38648.1 (NM_013620)	olfactory receptor 68 [<i>Mus musculus</i>]	315	258/314 (82%)	281/314 (89%)	e-122
gi 7305353 ref NP_0 38649.1 (NM_013621)	olfactory receptor 69 [<i>Mus musculus</i>]	316	255/314 (81%)	279/314 (88%)	e-120
gi 11908221 gb AAG4 1685.1 (AF133300)	MOR 3'Beta6 [<i>Mus musculus</i>]	316	238/311 (76%)	268/311 (85%)	e-115
gi 6912560 ref NP_0 36507.1 (NM_012375)	olfactory receptor, family 52, subfamily A, member 1 [<i>Homo sapiens</i>]	312	233/310 (75%)	263/310 (84%)	e-110

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 10D. In the ClustalW alignment of the NOV10 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and

can potentially be altered to a much broader extent without altering protein structure or function.

Table 10D. ClustalW Analysis of NOV10

5	1) Novel NOV10 (SEQ ID NO:50)																																																								
	2) gi 11991867 gb AAG42368.1 (AF289204) odorant receptor HOR3'beta5 [<i>Homo sapiens</i>] (SEQ ID NO:362)																																																								
	3) gi 7305351 ref NP_038648.1 (NM_013620) olfactory receptor 68 [<i>Mus musculus</i>] (SEQ ID NO:363)																																																								
10	4) gi 7305353 ref NP_038649.1 (NM_013621) olfactory receptor 69 [<i>Mus musculus</i>] (SEQ ID NO:364)																																																								
	5) gi 11908221 gb AAG41685.1 (AF133300) MOR 3'Beta6 [<i>Mus musculus</i>] (SEQ ID NO:365)																																																								
	6) gi 6912560 ref NP_036507.1 (NM_012375) olfactory receptor, family 52, subfamily A, member 1 [<i>Homo sapiens</i>] (SEQ ID NO:366)																																																								
15																																																									
	<table><tr><td></td><td></td><td>10</td><td>20</td><td>30</td><td>40</td><td>50</td><td>60</td></tr><tr><td>NOV10</td><td>1</td><td>MPTFNGSVFMPSAFIL</td><td>LGIPGLESVQCWIGIPFSAMY</td><td>LIGVIGNSLILVITIKYENSLHIP</td><td>60</td><td></td><td></td></tr><tr><td>gi 11991867 </td><td>1</td><td>MPTFNGSVFMPSAFIL</td><td>LGIPGLESVQCWIGIPFSAMY</td><td>LIGVIGNSLILVITIKYENSLHIP</td><td>60</td><td></td><td></td></tr><tr><td>gi 7305351 </td><td>1</td><td>MIKFNGSVFMPSVLT</td><td>VGIPGLESVQCWIGIPFCVMY</td><td>ITAMIGNSLILVITIKSEKSLHIP</td><td>60</td><td></td><td></td></tr><tr><td>gi 7305353 </td><td>1</td><td>MIKFNGSVFMPSVLT</td><td>VGIPGLESVQCWIGIPFCVMY</td><td>ITAMIGNSLILVITIKSEKSLHIP</td><td>60</td><td></td><td></td></tr><tr><td>gi 11908221 </td><td>1</td><td>MPHLNSTIFRPSVLT</td><td>LGIPGLESVQFWIGIPFCIMY</td><td>ITALLIGNSLILVITIKVERSLHEP</td><td>60</td><td></td><td></td></tr><tr><td>gi 6912560 </td><td>1</td><td>MSISNITIVYMPVLT</td><td>VGIPGLESVQCWIGIPFCAT</td><td>ITALLIGNSLILVITIKSEKSLHEP</td><td>60</td><td></td><td></td></tr></table>			10	20	30	40	50	60	NOV10	1	MPTFNGSVFMPSAFIL	LGIPGLESVQCWIGIPFSAMY	LIGVIGNSLILVITIKYENSLHIP	60			gi 11991867	1	MPTFNGSVFMPSAFIL	LGIPGLESVQCWIGIPFSAMY	LIGVIGNSLILVITIKYENSLHIP	60			gi 7305351	1	MIKFNGSVFMPSVLT	VGIPGLESVQCWIGIPFCVMY	ITAMIGNSLILVITIKSEKSLHIP	60			gi 7305353	1	MIKFNGSVFMPSVLT	VGIPGLESVQCWIGIPFCVMY	ITAMIGNSLILVITIKSEKSLHIP	60			gi 11908221	1	MPHLNSTIFRPSVLT	LGIPGLESVQFWIGIPFCIMY	ITALLIGNSLILVITIKVERSLHEP	60			gi 6912560	1	MSISNITIVYMPVLT	VGIPGLESVQCWIGIPFCAT	ITALLIGNSLILVITIKSEKSLHEP	60		
		10	20	30	40	50	60																																																		
NOV10	1	MPTFNGSVFMPSAFIL	LGIPGLESVQCWIGIPFSAMY	LIGVIGNSLILVITIKYENSLHIP	60																																																				
gi 11991867	1	MPTFNGSVFMPSAFIL	LGIPGLESVQCWIGIPFSAMY	LIGVIGNSLILVITIKYENSLHIP	60																																																				
gi 7305351	1	MIKFNGSVFMPSVLT	VGIPGLESVQCWIGIPFCVMY	ITAMIGNSLILVITIKSEKSLHIP	60																																																				
gi 7305353	1	MIKFNGSVFMPSVLT	VGIPGLESVQCWIGIPFCVMY	ITAMIGNSLILVITIKSEKSLHIP	60																																																				
gi 11908221	1	MPHLNSTIFRPSVLT	LGIPGLESVQFWIGIPFCIMY	ITALLIGNSLILVITIKVERSLHEP	60																																																				
gi 6912560	1	MSISNITIVYMPVLT	VGIPGLESVQCWIGIPFCAT	ITALLIGNSLILVITIKSEKSLHEP	60																																																				
20																																																									
	<table><tr><td></td><td></td><td>70</td><td>80</td><td>90</td><td>100</td><td>110</td><td>120</td></tr><tr><td>NOV10</td><td>61</td><td>MYIFLAMLAATDIAL</td><td>NTCILPKMLGIFWFHLP</td><td>ETISFDACL</td><td>FOMWLIHSFOAIESGILLAM</td><td>120</td><td></td></tr><tr><td>gi 11991867 </td><td>61</td><td>MYIFLAMLAATDIAL</td><td>NTCILPKMLGIFWFHLP</td><td>ETISFDACL</td><td>FOMWLIHSFOAIESGILLAM</td><td>120</td><td></td></tr><tr><td>gi 7305351 </td><td>61</td><td>MYIFLAAILAVTDIAL</td><td>NTCILPKMLGIFWFHMP</td><td>QISFDACL</td><td>LOMELIHSFOATESGILLAM</td><td>120</td><td></td></tr><tr><td>gi 7305353 </td><td>61</td><td>MYIFLAAILAVTDIAL</td><td>NTCILPKMLGIFWFHMP</td><td>QISFDACL</td><td>LOMELIHSFOATESGILLAM</td><td>120</td><td></td></tr><tr><td>gi 11908221 </td><td>61</td><td>MYIFLAMLGATDISL</td><td>STCILPKMLGIFWFHST</td><td>IYFDACL</td><td>LOMWLIHTFQIESGILFAM</td><td>120</td><td></td></tr><tr><td>gi 6912560 </td><td>61</td><td>LYIFLGM LGATDIAL</td><td>ASSIMPKMLGIFWFENV</td><td>PEIYFDS</td><td>CLLMWFHITLQIESGILVAM</td><td>120</td><td></td></tr></table>			70	80	90	100	110	120	NOV10	61	MYIFLAMLAATDIAL	NTCILPKMLGIFWFHLP	ETISFDACL	FOMWLIHSFOAIESGILLAM	120		gi 11991867	61	MYIFLAMLAATDIAL	NTCILPKMLGIFWFHLP	ETISFDACL	FOMWLIHSFOAIESGILLAM	120		gi 7305351	61	MYIFLAAILAVTDIAL	NTCILPKMLGIFWFHMP	QISFDACL	LOMELIHSFOATESGILLAM	120		gi 7305353	61	MYIFLAAILAVTDIAL	NTCILPKMLGIFWFHMP	QISFDACL	LOMELIHSFOATESGILLAM	120		gi 11908221	61	MYIFLAMLGATDISL	STCILPKMLGIFWFHST	IYFDACL	LOMWLIHTFQIESGILFAM	120		gi 6912560	61	LYIFLGM LGATDIAL	ASSIMPKMLGIFWFENV	PEIYFDS	CLLMWFHITLQIESGILVAM	120	
		70	80	90	100	110	120																																																		
NOV10	61	MYIFLAMLAATDIAL	NTCILPKMLGIFWFHLP	ETISFDACL	FOMWLIHSFOAIESGILLAM	120																																																			
gi 11991867	61	MYIFLAMLAATDIAL	NTCILPKMLGIFWFHLP	ETISFDACL	FOMWLIHSFOAIESGILLAM	120																																																			
gi 7305351	61	MYIFLAAILAVTDIAL	NTCILPKMLGIFWFHMP	QISFDACL	LOMELIHSFOATESGILLAM	120																																																			
gi 7305353	61	MYIFLAAILAVTDIAL	NTCILPKMLGIFWFHMP	QISFDACL	LOMELIHSFOATESGILLAM	120																																																			
gi 11908221	61	MYIFLAMLGATDISL	STCILPKMLGIFWFHST	IYFDACL	LOMWLIHTFQIESGILFAM	120																																																			
gi 6912560	61	LYIFLGM LGATDIAL	ASSIMPKMLGIFWFENV	PEIYFDS	CLLMWFHITLQIESGILVAM	120																																																			
25																																																									
	<table><tr><td></td><td></td><td>130</td><td>140</td><td>150</td><td>160</td><td>170</td><td>180</td></tr><tr><td>NOV10</td><td>121</td><td>ALDRYVAICIPLRHATIFS</td><td>QOFLTHICLGVT</td><td>LRAAILIIPSLGLIKCCKHYRTTVISHS</td><td>180</td><td></td><td></td></tr><tr><td>gi 11991867 </td><td>121</td><td>ALDRYVAICIPLRHATIFS</td><td>QOFLTHICLGVT</td><td>LRAAILIIPSLGLIKCCKHYRTTVISHS</td><td>180</td><td></td><td></td></tr><tr><td>gi 7305351 </td><td>121</td><td>ALDRYVAICNPLRHATIFS</td><td>POLTTCTGACALLRAFILV</td><td>SPSILLIKCRKLYRTTVISHS</td><td>180</td><td></td><td></td></tr><tr><td>gi 7305353 </td><td>121</td><td>ALDRYVAICNPLRHATIFS</td><td>POLTTCTGACALLRSLIT</td><td>TFPLILLIKFCKLYRTTVISHS</td><td>180</td><td></td><td></td></tr><tr><td>gi 11908221 </td><td>121</td><td>AMDRYVAICDPLRHATIFS</td><td>TOIGVGVTLRAALFVAP</td><td>CLFLIKCRKLYRTTVISHS</td><td>180</td><td></td><td></td></tr><tr><td>gi 6912560 </td><td>121</td><td>ALDRYVAICYPLRHANIF</td><td>THOLMIOIGTMVVLRAAIL</td><td>VAPCLVLIKCRFQFYHTTVISHS</td><td>180</td><td></td><td></td></tr></table>			130	140	150	160	170	180	NOV10	121	ALDRYVAICIPLRHATIFS	QOFLTHICLGVT	LRAAILIIPSLGLIKCCKHYRTTVISHS	180			gi 11991867	121	ALDRYVAICIPLRHATIFS	QOFLTHICLGVT	LRAAILIIPSLGLIKCCKHYRTTVISHS	180			gi 7305351	121	ALDRYVAICNPLRHATIFS	POLTTCTGACALLRAFILV	SPSILLIKCRKLYRTTVISHS	180			gi 7305353	121	ALDRYVAICNPLRHATIFS	POLTTCTGACALLRSLIT	TFPLILLIKFCKLYRTTVISHS	180			gi 11908221	121	AMDRYVAICDPLRHATIFS	TOIGVGVTLRAALFVAP	CLFLIKCRKLYRTTVISHS	180			gi 6912560	121	ALDRYVAICYPLRHANIF	THOLMIOIGTMVVLRAAIL	VAPCLVLIKCRFQFYHTTVISHS	180		
		130	140	150	160	170	180																																																		
NOV10	121	ALDRYVAICIPLRHATIFS	QOFLTHICLGVT	LRAAILIIPSLGLIKCCKHYRTTVISHS	180																																																				
gi 11991867	121	ALDRYVAICIPLRHATIFS	QOFLTHICLGVT	LRAAILIIPSLGLIKCCKHYRTTVISHS	180																																																				
gi 7305351	121	ALDRYVAICNPLRHATIFS	POLTTCTGACALLRAFILV	SPSILLIKCRKLYRTTVISHS	180																																																				
gi 7305353	121	ALDRYVAICNPLRHATIFS	POLTTCTGACALLRSLIT	TFPLILLIKFCKLYRTTVISHS	180																																																				
gi 11908221	121	AMDRYVAICDPLRHATIFS	TOIGVGVTLRAALFVAP	CLFLIKCRKLYRTTVISHS	180																																																				
gi 6912560	121	ALDRYVAICYPLRHANIF	THOLMIOIGTMVVLRAAIL	VAPCLVLIKCRFQFYHTTVISHS	180																																																				
30																																																									
	<table><tr><td></td><td></td><td>190</td><td>200</td><td>210</td><td>220</td><td>230</td><td>240</td></tr><tr><td>NOV10</td><td>181</td><td>YCEHMAIVKLATE</td><td>DIRVNKIYGLFVAFAILGFDI</td><td>IFITLSYVQIFITVFQLPQKEARFKA</td><td>240</td><td></td><td></td></tr><tr><td>gi 11991867 </td><td>181</td><td>YCEHMAIVKLATE</td><td>DIRVNKIYGLFVAFAILGFDI</td><td>IFITLSYVQIFITVFQLPQKEARFKA</td><td>240</td><td></td><td></td></tr><tr><td>gi 7305351 </td><td>181</td><td>YCEHMAIVKLAADIR</td><td>INIKICGLLVAFAILGFDIV</td><td>FITESYVRIFITVFQLPQKEARFKA</td><td>240</td><td></td><td></td></tr><tr><td>gi 7305353 </td><td>181</td><td>YCEHMAIVKLAADIR</td><td>INIKICGLLVAFAILGFDIV</td><td>FITESYVRIFITVFQLPQKEARFKA</td><td>240</td><td></td><td></td></tr><tr><td>gi 11908221 </td><td>181</td><td>YCEHMAIVKLAEDV</td><td>HVNKIYGLFVAFSILGLDII</td><td>IFITLSYTRIFITVFQLPQKEARLKA</td><td>240</td><td></td><td></td></tr><tr><td>gi 6912560 </td><td>181</td><td>YCEHMAIVKLAANV</td><td>QVNKIYGLFVAFVAGFDI</td><td>IFITLSYVQIFITVFQLPQKEARFKA</td><td>240</td><td></td><td></td></tr></table>			190	200	210	220	230	240	NOV10	181	YCEHMAIVKLATE	DIRVNKIYGLFVAFAILGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240			gi 11991867	181	YCEHMAIVKLATE	DIRVNKIYGLFVAFAILGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240			gi 7305351	181	YCEHMAIVKLAADIR	INIKICGLLVAFAILGFDIV	FITESYVRIFITVFQLPQKEARFKA	240			gi 7305353	181	YCEHMAIVKLAADIR	INIKICGLLVAFAILGFDIV	FITESYVRIFITVFQLPQKEARFKA	240			gi 11908221	181	YCEHMAIVKLAEDV	HVNKIYGLFVAFSILGLDII	IFITLSYTRIFITVFQLPQKEARLKA	240			gi 6912560	181	YCEHMAIVKLAANV	QVNKIYGLFVAFVAGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240		
		190	200	210	220	230	240																																																		
NOV10	181	YCEHMAIVKLATE	DIRVNKIYGLFVAFAILGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240																																																				
gi 11991867	181	YCEHMAIVKLATE	DIRVNKIYGLFVAFAILGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240																																																				
gi 7305351	181	YCEHMAIVKLAADIR	INIKICGLLVAFAILGFDIV	FITESYVRIFITVFQLPQKEARFKA	240																																																				
gi 7305353	181	YCEHMAIVKLAADIR	INIKICGLLVAFAILGFDIV	FITESYVRIFITVFQLPQKEARFKA	240																																																				
gi 11908221	181	YCEHMAIVKLAEDV	HVNKIYGLFVAFSILGLDII	IFITLSYTRIFITVFQLPQKEARLKA	240																																																				
gi 6912560	181	YCEHMAIVKLAANV	QVNKIYGLFVAFVAGFDI	IFITLSYVQIFITVFQLPQKEARFKA	240																																																				
35																																																									
	<table><tr><td></td><td></td><td>- 250</td><td>260</td><td>270</td><td>280</td><td>290</td><td>300</td></tr><tr><td>NOV10</td><td>241</td><td>FNTCIAHICVFLQFYLLA</td><td>FFSFFTHRFCSHIPPIHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGVKTK</td><td>300</td><td></td></tr><tr><td>gi 11991867 </td><td>241</td><td>FNTCIAHICVFLQFYLLA</td><td>FFSFFTHRFCSHIPPIHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGVKTK</td><td>300</td><td></td></tr><tr><td>gi 7305351 </td><td>241</td><td>FNTCIAHICVFLQFYLLA</td><td>FFSFFTHRFCAHIPPIVHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGVKTK</td><td>300</td><td></td></tr><tr><td>gi 7305353 </td><td>241</td><td>FNTCIAHICVFLQFYLLA</td><td>FFSFFTHRFCAHIPPIVHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGVKTK</td><td>300</td><td></td></tr><tr><td>gi 11908221 </td><td>241</td><td>FNTCVAHICVFLQFYLLA</td><td>FFSFFTHRFCHVPIHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGVKTK</td><td>300</td><td></td></tr><tr><td>gi 6912560 </td><td>241</td><td>FNTCIAHICVFLQFYLLA</td><td>FFSFFTHRFCSHIPPIHILLS</td><td>NLYLLVPPFLNP</td><td>IVYGA</td><td>AKTT</td><td>300</td></tr></table>			- 250	260	270	280	290	300	NOV10	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300		gi 11991867	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300		gi 7305351	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCAHIPPIVHILLS	NLYLLVPPFLNP	IVYGVKTK	300		gi 7305353	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCAHIPPIVHILLS	NLYLLVPPFLNP	IVYGVKTK	300		gi 11908221	241	FNTCVAHICVFLQFYLLA	FFSFFTHRFCHVPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300		gi 6912560	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGA	AKTT	300
		- 250	260	270	280	290	300																																																		
NOV10	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300																																																			
gi 11991867	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300																																																			
gi 7305351	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCAHIPPIVHILLS	NLYLLVPPFLNP	IVYGVKTK	300																																																			
gi 7305353	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCAHIPPIVHILLS	NLYLLVPPFLNP	IVYGVKTK	300																																																			
gi 11908221	241	FNTCVAHICVFLQFYLLA	FFSFFTHRFCHVPIHILLS	NLYLLVPPFLNP	IVYGVKTK	300																																																			
gi 6912560	241	FNTCIAHICVFLQFYLLA	FFSFFTHRFCSHIPPIHILLS	NLYLLVPPFLNP	IVYGA	AKTT	300																																																		
40																																																									
	<table><tr><td></td><td></td><td>310</td><td></td></tr><tr><td>NOV10</td><td>301</td><td>QIRDHIVKVFVFFKVT</td><td>316</td></tr></table>			310		NOV10	301	QIRDHIVKVFVFFKVT	316																																																
		310																																																							
NOV10	301	QIRDHIVKVFVFFKVT	316																																																						
45																																																									
50																																																									
55																																																									
60																																																									

5 gi|11991867| 301 QIRDHIVKVFVFFKVT 316
gi|7305351| 301 QIRDQVLKMLFSSKH- 315
gi|7305353| 301 QIRDQVLKMLFSSKPL 316
gi|11908221| 301 QIRDQVSKILYCNYSY 316
gi|6912560| 301 QIRIHVVKMFCS---- 312

10 Table 10E lists the domain description from DOMAIN analysis results against NOV10. This indicates that the NOV10 sequence has properties similar to those of other proteins known to contain this domain.

Table 10E Domain Analysis of NOV10

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 67.8 bits (164), Expect = 9e-13

15 NOV10: 43 GNSLILVVIKYENSLHIPMYIFLAMLAAITDIALNTCILPKMLGIFWPHLPEISFDACLFO 102
||| +++++ | | ||| | | + + | |
Sbjct: 1 GNLLVILVILRTKKLRTPNTIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV 60

20 NOV10: 103 MWLIHSFQAIESGILLAMALDRYVAICILRHATIFSQQFLTHIGLVTLRAAILIIPSL 162
| + | +++++ ||| + | + + + | + | + | + |
Sbjct: 61 GALFVVNGYASILLTAISIDRYLAIVHPLRYRRIRTPRRAKVLILLVWVLALLLSLPL 120

25 NOV10: 163 GLIKCCLKHYR-TTVISHSYCEHMAIVKLATEDIRVNKIYGLFVAFAILGF--DIIFITL 219
||| + | ++ + | + ++ | ||
Sbjct: 121 LFSWLRTVEEGNTTVCLIDFPEESVKRSYVL----LSTLVGFVLPPLLVLVCYTRILRTL 176

30 NOV10: 220 SYVQIFITVFQLPQKEARFKAFNTCIAHICVFLQF--YLLAFFSFFTHRFGSHIPPYIHI 277
+ | | + + | + | +
Sbjct: 177 RKRARSQRSLKRRSSSERKAAKMMLVVVVVFLCWLPHYHIVLLDLSLCLLSIWRVLPAL 236

NOV10: 278 LLSNLYLLVPPFLNPIVY 295
|++ | |||+|
Sbjct: 237 LITLWLAYVNSCLNPIIY 254

35 G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals.

Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

5 The disclosed NOV10 nucleic acid of the invention encoding a G-Protein Coupled Receptor -like protein includes the nucleic acid whose sequence is provided in Table 10A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 10A while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological
10 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example,
15 modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19 percent of the bases may be so changed.

20 The disclosed NOV10 protein of the invention includes the G-Protein Coupled Receptor-like protein whose sequence is provided in Table 10B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 10B while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological functions, or a functional fragment thereof.
25 In the mutant or variant protein, up to about 25 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

 The above disclosed information suggests that this G-Protein Coupled Receptor-like protein (NOV10) is a member of a "G-Protein Coupled Receptor family". Therefore, the
30 NOV10 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene

delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV10 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidolusian atrophy(DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV10 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV10 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV10 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV11

A disclosed NOV11 nucleic acid of 1014 nucleotides (also referred to as Curagen Accession No. CG55966-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 11A. An open reading frame was identified beginning with an ATG initiation

codon at nucleotides 2-4 and ending with a TGA codon at nucleotides 947-949. Putative untranslated regions upstream from the initiation codon and downstream of the termination codon are underlined in Table 11A. The start and stop codons are in bold letters.

Table 11A. NOV11 nucleotide sequence (SEQ ID NO:51).

AATGATTACTTCAGTAAGCCCTAGCACCAGCACGAATTCCTCTTCTCTCACTGGATTTTCTG
GCATGGAGCAGCAATACCCCTGGTTTTCCATCCCCTTCTCCTCAATCTATGCCATGGTGCTTTTG
GGCAATTGCATGGTTCTCCATGTGATATGGACTGAGCCAAGCCTGCACCAGCCTATGTTTTACTT
CCTGTCCATGCTGGCCCTCACTGACCTGTGCATGGGGCTGTCCACTGTGTACACAGTGTGGGA
TCCTGTGGCGGATCATTGAGAGATCAGCTTGGATTCTGCAATTGCCAGTCCTATTTTCATCCAT
GGTCTGTCTTCATGGAGTCTCTGTCTCTCACTATGGCCTTTGACCGGTACATTGCAATTTG
CAATCCACTACGTTATTCCTCCATCCTGACTAATTCAGAATTATCAAAATTGGGCTCACTATAA
TAGGTAGGAGTTTTTCTTTATTACACCCCCATCATCTGTCTGAAATTTTTTAATTACTGTCAAT
TTCCACATCCTTTCTCACTCTTTCTGCCTGCACCAGGATCTTCTCCGCTTAGCCTGTTTCAACAT
CCGATTCAATAGTTACTATGCCCTGATGCTGGTTATTTGCATACTGTTGTTGGATGCTATACTCA
TCCTTTTCTCCTACATCCTGATTCTTAAGTCAGTCTCGCAGTTGCCTCTCAGGAAGAGAGGCAT
AAATTATTTTCAAGCTGCATCTCCACATCTGTGCTGTCTTGTGTTCTACATCCCTATCATTAG
CCTCACAATGGTGCACCGTTTTGGCAAGCACCTTTCCCCCGTGGCCCACGTTCTCATTGGCAACA
TCTACATCCTTTTCCCACCTTTAATGAATCCCATCATCTACAGTGTCAAGACCCAACAGATTTCAT
ACCAGAATGCTTAGACTCTTTTCTCTGAAAAGATATTGAGAGATATTGAGATGTATTGCCTAAAA
AAAAGAAAGAAAAGCAGCAACAATAATAACAAAAATCA

The disclosed NOV11 polypeptide (SEQ ID NO:52) encoded by SEQ ID NO:51 has 315 amino acid residues and is presented in Table 11B using the one-letter amino acid code.

Table 11B. Encoded NOV11 protein sequence (SEQ ID NO:52).

MITVSPSTSTNSSFLLTGFSGMEQQYPWFSSIPFSSIAMVLLGNMVLHVIWTEPSLHQPMFY
FLSMLALTDLCMGLSTVYTVLGILWRIIREISLDSCIAQSYFIHGLSFMESSVLLTMAFDRIYA
ICNPLRYSSILTNSRIKIGLTIIGRSFFFITPPIICLKFFNYCHFHLSHSFCLHQDLLRLAC
SDIRFNSYYALMLVICILLDAILILFSYILILKSVLAVASQEERHKLFPQTCISHICAVLVFYI
PIISLTMVHRFGKHLSPVAHVLIIGNIYILFPLMNPPIYSVKTQIHTMRMLRFLSKRY

A search of sequence databases reveals that the NOV11 amino acid sequence has 165 of 302 amino acid residues (54%) identical to, and 222 of 302 amino acid residues (73%) similar to, the 311 amino acid residue ptnr: SPTREMBL-ACC:Q9WVN4 protein from *Mus musculus* (Mouse) MOR 5'BETA1 ($E = 7.0e^{-88}$).

The disclosed NOV11 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 11C.

Table 11C. BLAST results for NOV11

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 11991863 gb AAG4 2364.1 (AF289204)	odorant receptor HOR3'beta1 [<i>Homo sapiens</i>]	321	315/315 (100%)	315/315 (100%)	e-139

gi 11908218 gb AAG41683.1 (AF137396)	HOR5'Beta5 [<i>Homo sapiens</i>]	312	165/307 (53%)	231/307 (74%)	4e-78
gi 17456753 ref XP_061614.1 (XM_061614)	similar to MOR 3Beta4 (H. sapiens) [<i>Homo sapiens</i>]	315	163/307 (53%)	223/307 (72%)	1e-77
gi 7305345 ref NP_038645.1 (NM_013617)	olfactory receptor 65 [<i>Mus musculus</i>]	307	164/305 (53%)	223/305 (72%)	5e-77
gi 17456767 ref XP_061618.1 (XM_061618)	similar to prostate specific G-protein coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	879	162/303 (53%)	226/303 (74%)	2e-76

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 11D. In the ClustalW alignment of the NOV11.protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 11D. ClustalW Analysis of NOV11

- 1) Novel NOV11 (SEQ ID NO:52)
- 2) gi|11991863|gb|AAG42364.1| (AF289204) odorant receptor HOR3'beta1 [*Homo sapiens*] (SEQ ID NO:367)
- 3) gi|11908218|gb|AAG41683.1| (AF137396) HOR5'Beta5 [*Homo sapiens*] (SEQ ID NO:368)
- 4) gi|17456753|ref|XP_061614.1| (XM_061614) similar to MOR 3Beta4 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:369)
- 5) gi|7305345|ref|NP_038645.1| (NM_013617) olfactory receptor 65 [*Mus musculus*] (SEQ ID NO:370)
- 6) gi|17456767|ref|XP_061618.1| (XM_061618) similar to prostate specific G-protein coupled receptor (H. sapiens) [*Homo sapiens*] (SEQ ID NO:371)

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

			610	620	630	640	650	660	
5	NOV11	34	FSSTIYAMVLLGNCMVLLHVIWTEPSLHOPMEYFLSMLALTDLCMGLSTVYTVLGGILWRIIR	93					
	gi 11991863	40	FSSTIYAMVLLGNCMVLLHVIWTEPSLHOPMEYFLSMLALTDLCMGLSTVYTVLGGILWRIIR	99					
	gi 11908218	29	FLFMYISTLLFCNGTLLLLIKEDHNLHEPMEYFLAMLAATDILGIALTTMPTVLGGVLWIDHR	88					
	gi 17456753	33	FCLAYLVAFMGNVTILSVIWIWESSLHOPMEYFISILAVNDLGMSLSTLPTMLAVLWLDAP	92					
	gi 7305345	29	FFATYILSVLLGNGTLLYLKDDHNLHEPMEYFLAMLAATDILTVLTMTVMAYLVWNHR	88					
	gi 17456767	601	LCFMYLVSTFCNCTILFTIKTERSLHEPMEYFLSMLALTDLGSLCTLPTVLGGIFWVGAR	660					
10			670	680	690	700	710	720	
	NOV11	94	EISLDSCIAQSYFIHGLSFMESSVLLTMAFDRIAIAICNPLRYSSILTNSRIITKIGLTIIG	153					
	gi 11991863	100	EISLDSCIAQSYFIHGLSFMESSVLLTMAFDRIAIAICNPLRYSSILTNSRIITKIGLTIIG	159					
	gi 11908218	89	EIGSAACFSAQYFIHLSLFLSGILLAMAMDRFIAICNPLRYTSVLNTNRVVKIGLGVLM	148					
15	gi 17456753	93	EIQASACFQAQLFEIHTFTLESSVLLAMAFDRFVAICNPLHYPIITLNSVICKIGLACLIL	152					
	gi 7305345	89	EIRHGACFLQAVIIHSLSLVESGVLLAMSYDRFVAICTPLHYNSILTNSRVTAIGLGVVL	148					
	gi 17456767	661	EISHDACFQAQLFEIHCFSLESSVLLSMAFDRFVAICNPLHYVTSILTNTVTIGRIGLVSIG	720					
20			730	740	750	760	770	780	
	NOV11	154	RSFFFTITPPIICLKFFFNCHFHILSHSFCLHQDLLRLACSDIRFNSYYALMLVICITILLD	213					
	gi 11991863	160	RSFFFTITPPIICLKFFFNCHFHILSHSFCLHQDLLRLACSDIRFNSYYALMLVICITILLD	219					
	gi 11908218	149	RGFVSVVPPIRPLFYFFLYCHSHVLSHAFCLHQDVILKACADTTFNRLYPAVILVVFIFVLD	208					
	gi 17456753	153	RSLGVVLPPTPLLRLHYHYCHGNALSHAFCLHQDVLRLSCTDARTNSIYGLCVVIATLGVLD	212					
25	gi 7305345	149	RGFLSLVPPILPLFWFSYCRSHVLSHAFCLHQDVMMKLACADITFNRIYPPVILVALTFFLD	208					
	gi 17456767	721	RSVALITPLPLPMLKRFPCYCGSPVLSHSYCLHQDVMMKLACADMKANSIYGMFVLVSTVGLD	780					
30			790	800	810	820	830	840	
	NOV11	214	AILILFSYVILILKSVLAVASQEERKHLFOTCHSHICAVLVFYIPIITSLTMVHRFGKHLSP	273					
	gi 11991863	220	AILILFSYVILILKSVLAVASQEERKHLFOTCHSHICAVLVFYIPIITSLTMVHRFGKHLSP	279					
	gi 11908218	209	YLILFISYVILIKTVLSTASREERAKALITCVSHICCVLVFYITVIGLSLTHRFGKVPH	268					
	gi 17456753	213	SIFILLSYVILINTVLDASREEQLKALNTCVSHICVILFFVEVTIGSMVHRFGKHLSP	272					
	gi 7305345	209	ALILFISYVILIKTVMGSTASGEERAKALNTCVSHISCVLVFYITVIGLFTTHRFGKNAPH	268					
35	gi 17456767	781	SLLILFSYVALILRTVLSTASRAERFKALNTCVSHICAVILFYTEMIGLSVTHRFGKQAPH	840					
40			850	860	870	880			
	NOV11	274	VAHVLLIGNIYLLFPPLMNPITYSVKTKQIHTRMRLRFLSKRY--	315					
	gi 11991863	280	VAHVLLIGNIYLLFPPLMNPITYSVKTKQIHTRMRLRFLSKRY--	321					
	gi 11908218	269	IVHLLMSYAYLLFPPLMNPITYSVKTKQIQAHLHLFTTHRIGT	312					
	gi 17456753	273	IVHLLMADIYLLLPVLPNPVSVRTRKQIRLGILHKKFVIRRRF-	315					
	gi 7305345	269	VVHLLMSYVYLLFPPLMNPITYSKTKQIQRSVLLHLLSV----	307					
45	gi 17456767	841	LVQVVMGFMYLLFPFVMPNPVSVKTKQIRDRVTHAFCY-----	879					

Table 11E lists the domain description from DOMAIN analysis results against NOV11. This indicates that the NOV11 sequence has properties similar to those of other proteins known to contain this domain.

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 71.2 bits (173), Expect = 8e-14

55

NOV11:	44	GNCMVLVHVIWTEPSLHQPMFYFLSMLALTDLCMGLSTVYTVLGILWRIIREISLDSCIAQ	103
		+ + + +	
Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
NOV11:	104	SYFIHGLSFMESSVLLTMAFDRIAICNPLRYSSILTNSRIIKIGLTIIGRSFFFITPPI	163

5
 10
 15

```

      +   +|  ++ |||+|| +||| | | | + | +   ||+
Sbjct: 61  GALFVVNGYASILLTIAISIDRYLAIVHPLRYRRIRTPRRAKVLILLVWVLALLSLPPL 120
NOV11: 164  ICLKFFNYCHFHLSHSFCLHQDLLRLACSDIRFNSYYALMLVICILLDDAILILFSYIL 223
      + |      + + ||      +      | | + + + | ++| |
Sbjct: 121  L---FSWLRTVEEGNTTVCCLIDF-----PEESVKRSYVLLSTLVGFVLPPLLVILVCYTR 171
NOV11: 224  ILKSVLAVA-----SQEERHKLFQTCISHICAVLVF--YIPIISLTMVHRFGKHL 272
      ||+++ |      | ||      + + || + | ++ | +
Sbjct: 172  ILRTLKRKRARSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLDLCLLSIWRV 231
NOV11: 273  PVAHVLIIGNIYILFPPLMNPIIY 295
      +||      +|||
Sbjct: 232  LPTALLITLWLAYVNSCLNPIIY 254
  
```

20

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

25
 30

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

35
 40

The disclosed NOV11 nucleic acid of the invention encoding a G-Protein Coupled Receptor-like protein includes the nucleic acid whose sequence is provided in Table 11A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 11A while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example,

modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

5 The disclosed NOV11 protein of the invention includes the G-Protein Coupled Receptor-like protein whose sequence is provided in Table 11B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 11B while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological functions, or a functional fragment thereof.

10 In the mutant or variant protein, up to about 47 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above disclosed information suggests that this G-Protein Coupled Receptor-like protein (NOV11) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV11 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV11 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological

disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluisian atrophy(DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other pathologies and disorders.

NOV11 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV11 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV12

A disclosed NOV12 nucleic acid of 1067 nucleotides (also referred to as Curagen Accession No. CG56003-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 12A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 15-17 and ending with a TGA codon at nucleotides 1023-1025. The untranslated regions are underlined and the start and stop codons are in bold letters in Table 12A.

Table 12A. NOV12 nucleotide sequence (SEQ ID NO:53).

AAAACCTGACATAAATGAACAACAATACAACATGTATTCAACCATCTATGATCTCTCCATGGCTTTACCAA
 TCATTTACATCCTCCTTTGTATTGTTGGTGTTTTGGAAACACTCTCTCAATGGATATTTTAAACAAAA
 TAGGTAAAAAACATCAACGCACATCTACCTGTCACACCTTGTGACTGCAAACTTACTTGTGTGCAGTGCCA
 TGCCTTTTCATGAGTATCTATTTCTGAAAGGTTTCCAATGGGAATATCAATCTGCTCAATGCAGAGTGGTCA
 ATTTTCTGGGAACCTCTATCCATGCAATGCAAGTATGTTTGTGAGTCTCTTAATTTAAGTTGGATTGCCATAA
 GCCGCTATGCTACCTTAATGCAAAAGGATTCTCGCAAGAGACTACTTCATGCTATGAGAAAAATTTTATG
 GCCATTTACTGAAAAAATTCGCCAGCCCAACTTGTCTAGAAAACATATGCATTTACATATGGGGAGTTGTAC
 TGGGCATAATCATTCCAGTTACCGTATACTACTCAGTCATAGAGGCTACAGAAGGAGAAGAGAGCCTATGCT
 ACAATCGGCAGATGGAACCTAGGAGCCATGATCTCTCAGATTGCAGGTCTCATTGGAACCACATTTATTGGAT
 TTTCTTTTGTAGTAGTACTAACATCATACTACTCTTTTGTAAAGCCATCTGAGAAAAATAAGAACCTGTACGT
 CCATTATGGAGAAAGATTGACTTACAGTTCTGTGAAAAGACATCTTTGGTCATCCAGATTCTACTAATAG
 TTTGCTTCTTCTTATAGTATTTTAAACCCATTTTATGTTCTACACCAAAGAGATAACTGTGAGCAAT
 TGAATATTTAATAGAAACAAAAACATTCTCAGCTGTCTTGTCTCGGCCAGAAGTAGCACAGCCCCATTA
 TATTTCTTTTATTAGATAAAACATTCAAGAAGACACTATATAATCTCTTTACAAAGTCTAATTCAGCACATA
 TGAATCATATGGT**TGA**CTTTTGAATGGAAAACCCCAATATTAAGAAAAGCATTCAT

The disclosed NOV12 polypeptide (SEQ ID NO 54) encoded by SEQ ID NO:53 has 336 amino acid residues and is presented in Table 12B using the one-letter amino acid code.

Table 12B. Encoded NOV12 protein sequence (SEQ ID NO:54).

MNNNTTCIQPSMISSMALPIIYILLCIVGVFGNTLSQWIFLT'KIGKKTSTHIYLSHLVTANLLV
CSAMPFMSIYFLKGFQWEYQSAQCRVVNFLGTLSMHASMFVSLILSWIAISRYATLMQKDSQ
ETTSCYEKIFYGHLLKKFRQPNFARKLCIYIWGVVLGIIIPVTVYYSVIEATEGEESLCYNRQM
ELGAMISQIAGLIGTTFIGFSFLVVLTSYYSFVSHLRKIRTCTSIMKDLTYSSVKRHLVLIQI
LLIVCFLPYSIFKPIFYVLHQDNCQQLNYLIETKNILTCLASARSSTDPIIFLLLDKTFKKTLYNLFT
KNSAHMQSYG

A search of sequence databases reveals that the NOV12 amino acid sequence has 52 of 179 amino acid residues (29%) identical to, and 86 of 179 amino acid residues (48%) similar to, the 339 amino acid residue ptnr: SWISSPROT-ACC:Q13304 protein from *Homo sapiens* Putative G Protein-Coupled Receptor GPR17 (R12) ($E = 1.6e^{-22}$).

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV12 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 12C.

Table 12C. BLAST results for NOV12

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 18201870 ref NP_543007.1 (NM_080817)	G protein-coupled receptor 82 [<i>Homo sapiens</i>]	336	336/336 (100%)	336/336 (100%)	e-170
gi 4885301 ref NP_05282.1 (NM_005291)	G protein-coupled receptor 17 [<i>Homo sapiens</i>]	367	85/322 (26%)	144/322 (44%)	6e-21
gi 17462169 ref XP_002705.4 (XM_002705)	G protein-coupled receptor 17 [<i>Homo sapiens</i>]	339	85/322 (26%)	144/322 (44%)	2e-20
gi 2695876 emb CAB08108.1 (Z94155)	P2Y-like G-protein coupled receptor [<i>Homo sapiens</i>]	298	80/302 (26%)	135/302 (44%)	3e-18
gi 5757634 gb AAD50531.1 AF039686_1 (AF039686)	G-protein coupled receptor GPR34 [<i>Homo sapiens</i>]	381	77/323 (23%)	152/323 (46%)	4e-18

5	gi 18201870	197	MISQIAGLIGTTFIGFSFIVVLTSSVSFVSHLRKIRTCTSIMKDLTYSSVKRHLLVIOI	256
	gi 4885301	219	HHALVS--LAVAFT-FPEITTVTCYLLIIRSLRQ-----LRVEKRLKTKAVR-MIAIVLA	270
	gi 17462169	191	HHALVS--LAVAFT-FPEITTVTCYLLIIRSLRQ-----LRVEKRLKTKAVR-MIAIVLA	242
	gi 2695876	150	HHALVS--LAVAFT-FPEITTVTCYLLIIRSLRQ-----LRVEKRLKTKAVR-MIAIVLA	201
	gi 5757634	214	KGEATFNFLVVMFWLIELLILSLVIKIGKNLLRISKRRSKFPNSGKYATTARNSFIVLI	273
10	NOV12	257	LLIVCFLPYSIFKPIFIVLHQRD--NQQLNYLIETKNILTCLASARSSTDPIIFLLLDK	314
	gi 18201870	257	LLIVCFLPYSIFKPIFIVLHQRD--NQQLNYLIETKNILTCLASARSSTDPIIFLLLDK	314
	gi 4885301	271	IFLVCFVPYHVNRSVYVLIHYRSHGASCATORILALANRITSCLTSLNGALDPIMYFFVAE	330
	gi 17462169	243	IFLVCFVPYHVNRSVYVLIHYRSHGASCATORILALANRITSCLTSLNGALDPIMYFFVAE	302
	gi 2695876	202	IFLVCFVPYHVNRSVYVLIHYRSHGASCATORILALANRITSCLTSLNGALDPIMYFFVAE	261
15	gi 5757634	274	IFLVCFVPYHAFRFIVLSSQLNVS-SCYWKETVHKTNELMLVLSSFNSCLDPVYFILMSS	332
20	NOV12	315	TEKKTLYNLFT---K-----SNSAHMQSYG---	336
	gi 18201870	315	TEKKTLYNLFT---K-----SNSAHMQSYG---	336
	gi 4885301	331	KFRHALCNLLCGKRLKGPPP-----SFEGKTNESSLSAKSEL-	367
	gi 17462169	303	KFRHALCNLLCGKRLKGPPP-----SFEGKTNESSLSAKSEL-	339
	gi 2695876	262	KFRHALCNLLCGKRLKGPPP-----SFEGKTNESSLSAKSEL-	298
25	gi 5757634	333	NIRKIMCOLLFR-RFOGPPSRSESTSEFKPGYSLHDTSVAVKIQSSSKST	381

Table 12E lists the domain description from DOMAIN analysis results against NOV12. This indicates that the NOV12 sequence has properties similar to those of other proteins known to contain this domain.

Table 12E Domain Analysis of NOV12

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 99.6% aligned
Score = 82.0 bits (201), Expect = 5e-17

30	NOV12:	32	GNTLSQWIFLTKIGKKTSTHIYLSHLVTANLLVCSAMPFMSIYFLKGFQWEYQSAQCRVV	91
	Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
35	NOV12:	92	NFLGTLMSHSMFVSLILSWIAISRYATLMQKDSSQETTSCYEKIFYGHLLKKFRQPNF	151
	Sbjct:	61	GALFVVNGYASIL---LLTAISIDRYLA-----IVHPLRYRRIRTPRR	100
40	NOV12:	152	ARKLCIYIWGVVLGIIIPVTVVYYSVIEATEGEESLCYNRMELGAMISQIAGLIGTTFIG	211
	Sbjct:	101	AKVLILLVVWLALLLSLPLLFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFV-	159
45	NOV12:	212	FSFLVVLTSYYSFVSHLRK-IRTCTSIMKDLTYSSVKRHLLVIQILLIVCFLPYSIFKP	270
	Sbjct:	160	LPLLVLVCYTRILRTLKRKARSQRSKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLL	219
50	NOV12:	271	IFYVLHQDNCQQLNYLIETKNILTCLASARSSTDPII	308
	Sbjct:	220	LDLCLLSIWRVLPT----ALLITLWLAYVNSCLNPPII	253

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven

transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals.

Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV12 nucleic acid of the invention encoding a G-Protein Coupled Receptor-like protein includes the nucleic acid whose sequence is provided in Table 12A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 12A while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

The disclosed NOV12 protein of the invention includes the G-Protein Coupled Receptor-like protein whose sequence is provided in Table 12B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 12B while still encoding a protein that maintains its G-Protein Coupled Receptor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 77 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor-like protein (NOV12) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV12 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug

targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

5 The NOV12 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to
10 starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and
15 Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease
20 or Gilles de la Tourette syndrome, and/or other pathologies and disorders.

NOV12 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV12 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
25 NOVX Antibodies" section below. The disclosed NOV12 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

30 NOV13

NOV13 includes three novel G-Protein Coupled Receptor -like proteins disclosed below. The disclosed sequences have been named NOV13a and NOV13b.

NOV13a

A disclosed NOV13a nucleic acid of 961 nucleotides (also referred to as CG56075-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 13A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 12-14 and ending with a TGA codon at nucleotides 936-938. The start and stop codons are shown in bold in Table 13A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 13A. NOV13a nucleotide sequence (SEQ ID NO:55).

GACAACAAACTATGAGACAGATAAATCAGACACAAGTGACAGAATTCCTCCTTCTGGGACTCTCTGATGGGC
CACACACCGAGCAGCTGCTATTTATCGTATTATTGGGTGTCTACCTGGTCACTGTGCTTGGAAATCTGCTTC
TAATCTCCCTTGTTTCATGTTGACTCCCAACTTCACACACCCATGTATTTTTTCTCTGCAACTTGTCTCTGG
CTGACCTCTGTTTCTCTACCAACATAGTTCTCAGGCACTAGTCCACCTGCTTCCAGAAAGAAGGTCATTG
CATTCACACTTTGCGCAGCTCGACTTCTCTTTTCTCCTCATTTTGGGTGTACCCAGTGCGCCCTTCTGCAG
TGATGTCCTATGATCGCTATGTTGCAATCTGCAATCCTCTGCGTTACCCTAACATCATGACCTGGAAAGTGT
GTGTCCAGCTGGCAACAGGATCATGGACCAGTGGCATTCTGGTGTCTGTGGTAGACACCACCTTCACACTGA
GGCTACCTTACCGAGGCAGTAACAGCATTGCTCATTCTTTTGTGAGGCCCTGCACTATGTAGCTTAGCAT
CCACAGACACCCATGCATCAGAGATGGCCATTTTCTTACGGGGTGTGATTCTCCTCATACCTGTTTTTC
TGATTCTGGTATCCTATGGCCGTATCATAGTAAGTGTGGTCAAGATGAAGTCAACTGTGGGGAGTCTCAAGG
CATTTTCTACCTGTGGCTCCCACTCATGGTGGTCATACTTTTTTATGGATCAGCAATTATCACTTACATGA
CACCCAAGTCTTCCAAACAGCAGGAAAAATCGGTGTCTGTTTCTATGCAATAGTGACTCCCATGCTTAATC
CCCTCATCTATAGCCTGAGAAACAAGGATGTGAAGGCAGCTCTGAGGAAAGTAGCCACAAGGAATTTCCCAT
GAAGGCTTGAATCTCACACTGACA

The disclosed NOV13a polypeptide (SEQ ID NO:56) encoded by SEQ ID NO:55 has 308 amino acid residues and is presented in Table 13B using the one-letter amino acid code.

Table 13B. Encoded NOV13a protein sequence (SEQ ID NO:56).

MRQINQTVTEFLLLGLSDGPHTQQLLFIVLLGVYLVTVLGNLLLSLVHVDSQLHTPMYFFLC
NLSLADLCFSTNIVPQALVHLLSRKKVIAFTLCAARLLFFLI FGCTQCALLAVMSYDRYVAICN
PLRYPNIMTWKVCVQLATGSWTSGILVSVDTTFTLRLPYRGSNSIAHFFCEAPALLILASTDT
HASEMAIFLTGVVILLIPVFLILVSYGRIIVTVVKMKSTVGSLKAFSTCGSHLMVVILFYGSAI
ITYMTPKSSKQKEKSVSVFYAIVTPMLNPLIYSLRNKDVKAALRKVATRNFP

A search of sequence databases reveals that the NOV13a amino acid sequence has 216 of 217 amino acid residues (99%) identical to, and 217 of 217 amino acid residues (100%) similar to, the 217 amino acid residue ptnr: SPTREMBL-ACC:O95224 protein from *Homo sapiens* (Human) (Olfactory Receptor) ($E = 2.2e^{-109}$).

NOV13b

A disclosed NOV13b nucleic acid of 961 nucleotides (also referred to as CG56021-02) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 13C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 12-14 and ending with a TGA codon at nucleotides 936-938. A putative untranslated region upstream

from the initiation codon is underlined in Table 13C. The start and stop codons are in bold letters.

Table 13C. NOV13b nucleotide sequence (SEQ ID NO:57).

GACAACAACTATGAGACAGATAAATCAGACACAAGTGACAGAATTCCTCCTTCTGGGACTCTGTGATGGGC
CACACACCGAGCAGCTGCTATTTATCGTATTATTTGGGTGTCTACCTGGTCACTGTGCTTGGAAATCTGCTTC
TAATCTCCCTTGTTTCATGTTGACTCCCACTTCACACACCCATGTATTTTTTCTCTGCAACTGTCTCTGG
CTGACCTCTGTTTCTCTACCAACATAGTTCCTCAGGCACTAATCCACCTGCTTTCCAGAAAGAAGGTCATTG
CAATTCACACTTTGCGCAGCTCGACTTCTTTTTCTCATTCTTTGGGTGTACCCAGTGCGCCCTTCTTGCAG
TGATGTCCTATGATCGCTATGTTGCAATCTGCAATCCTCTGCGTTACCCTAACATCATGACCTGGAAAGTGT
GTGTCCAGCTGGCAACAGGATCATGGACAGTGGCATTCTGGTGTCTGTGGTAGACACCACCTTCACACTGA
GGCTACCCTACCGAGGCAGTAACAGCATTGCTCATTCTTTTGTGAGGCCCTGCACTATTGATCTTAGCAT
CCACAGACACCCATGCATCAGAGATGGCCATTTTCTTATGGGGGTTGTGATTCTCTCATACCTGTTTTTC
TGATTCTGGTATCCTATGGCCGTATCATAGTAAGTGTGGTCAAGATGAAGTCAACTGTGGGGAGTCTCAAGG
CATTTCTACCTGTGGCTCCACCTCATGGTGGTCATACTTTTTATGGATCAGCAATTATCACTTGCATGA
CACCAAGTCTTCCAAACAGCAGGAAAAATCGGTGTCTGTTTTCTATGCAATAGTGACTCCCATGCTTAATC
CCCTCATCTATAGCCTGAGAAACAAGGATGTGAAGGCAGCTCTGAGGAAAGTAGCCACAAGGAATTTCCCAT
GAAGGCTTGAATCTCACTGACA

In a search of public sequence databases, the NOV13b nucleic acid sequence has 648
of 653 bases (99%) identical to a gb:GENBANK-ID:AF065876|acc:AF065876.1 mRNA from
Homo sapiens (olfactory receptor (OR2D2) gene, partial cds) ($E = 2.8e^{-139}$).

The disclosed NOV13b polypeptide (SEQ ID NO:58) encoded by SEQ ID NO:57 has
308 amino acid residues and is presented in Table 13D using the one-letter amino acid code.
Signal P, Psort and/or Hydropathy results predict that NOV13b has a signal peptide and is
likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively,
NOV13b may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic
reticulum (membrane) with a certainty of 0.3000, or in the microbody (peroxisome) with a
certainty of 0.3000. The most likely cleavage site for NOV13b is between positions 53 and 54:
VDS-QL.

Table 13D. Encoded NOV13b protein sequence (SEQ ID NO:58).

MRQINQTVTEFLLGLCDGPHTQLLFIVLLGVYLVTVLGNLLISLVHVDSQLHTPMYFFLCNLSLADLC
FSTNIVPQALIHLLSRKKVIAFTLCAARLLFFLIFGCTQCALLAVMSYDRYVAICNPLRYPNIMTWKVCVQL
ATGSWTSGILVSVVDTTFLRLPYRGSNSIAHFFCEAPALLILASTDTHASEMAIFLMGVVILLIPVFLILV
SYGRIIVTVVKMKSTVGSLKAFSTCGSHLMVILFYGSAIITCMTPKSSKQEKSVSVFYAIVTPMLNPLIY
SLRNKDVKAALRKVATRNFP

A search of sequence databases reveals that the NOV13 amino acid sequence has 52 of
179 amino acid residues (29%) identical to, and 86 of 179 amino acid residues (48%) similar
to, the 339 amino acid residue ptmr: SWISSPROT-ACC:Q13304 protein from *Homo sapiens*
Putative G Protein-Coupled Receptor GPR17 (R12) ($E = 3.3e^{-157}$).

NOV13b is predicted to be expressed in at least Apical microvilli of the retinal
pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines,

corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV13a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 13E.

Table 13E. BLAST results for NOV13a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14423807 sp Q9H210 O2D2_HUMAN	OLFACTORY RECEPTOR 2D2 (OLFACTORY RECEPTOR 11-610) (OR11-610) (HB2)	308	307/308 (99%)	308/308 (99%)	e-148
gi 17461460 ref XP_062286.1 (XM_062286)	similar to hB2 olfactory receptor (H. sapiens) [<i>Homo sapiens</i>]	308	308/308 (100%)	308/308 (100%)	e-148
gi 12007409 gb AAG45183.1 (AF321233)	B2 olfactory receptor [<i>Mus musculus</i>]	314	261/305 (85%)	278/305 (90%)	e-127
gi 3831619 gb AAC70020.1 (AF065876)	olfactory receptor [<i>Homo sapiens</i>]	217	216/217 (99%)	217/217 (99%)	e-100
gi 15293767 gb AAK95076.1 (AF399591)	olfactory receptor [<i>Homo sapiens</i>]	214	213/214 (99%)	214/214 (99%)	e-100

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 13F. In the ClustalW alignment of the NOV13 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.


```

NOV13a      301  KVATRNFP----- 308
NOV13b      301  KVATRNFP----- 308
gi|14423807| 301  KVATRNFP----- 308
gi|17461460| 301  KVATRNFP----- 308
gi|12007409| 300  KVAMKNFSSRLRITH 314
gi|3831619|  217  ----- 217
gi|15293767| 214  ----- 214

```

Table 13G lists the domain description from DOMAIN analysis results against NOV13. This indicates that the NOV13 sequence has properties similar to those of other proteins known to contain this domain.

Table 13G Domain Analysis of NOV13

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
 CD-Length = 254 residues, 94.9% aligned
 Score = 93.2 bits (230), Expect = 2e-20

```

NOV13:  54  QLHTPMYFFLCNLSLADLCFSTNIVPQALVHLLSRKKVIAFTLCAARLLFFLIFGCTQCA 113
      +| || || ||++|| | + | || ++ | || ++ |
Sbjct:  14  KLRTPTNIFLLNLAVADLLFLLTLPWALYYLVGGDWVFGDALCKLVGALFVVNGYASIL 73

NOV13: 114  LLAVMSYDRYVAICNPLRYPNIMTWKVCVQLATGSWTSGILVSVDTTFTLRLPYRGSNS 173
      || +| |||+|| +||| | | + | | +|+| + | +
Sbjct:  74  LLTAISIDRYLAIVHPLRYRRIPTPRRAKVLILLVWVLALLSLPPLLFSLWRTVEEGNT 133

NOV13: 174  IAHFFC-----EAPALLILASTDTHASEMAIFLTGVVILLIPVFLILVSYGRIIVTVVKM 228
      + ++|++ + + | | | + | + |
Sbjct: 134  TVCLIDFPESVKRSYVLLSTLVGFVLPLLVILV-----CYTRILRTLKRK 179

NOV13: 229  KSTVGS�K-----AFSTCGSHLMVVILFYGSAIITYMTPKSSKQEKSVSVFYAI- 278
      + ||| | ++ | + + | + + + + |
Sbjct: 180  ARSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSLCLLSIWRVLP TALLIT 239

NOV13: 279  -----VTPMLNPLIY 288
      | |||+||
Sbjct: 240  LWLAYVNSCLNPIIY 254

```

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals.

5 Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV13 nucleic acid of the invention encoding a G-Protein Coupled
10 Receptor -like protein includes the nucleic acid whose sequence is provided in Table 13A, 14C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 13A, or 14C while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
15 nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones
20 are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

25 The disclosed NOV13 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 13B, or 14D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 13B, or 14D while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a
30 functional fragment thereof. In the mutant or variant protein, up to about 15 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV13) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV13 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV13 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV13 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV13 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV13 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in

understanding of pathology of the disease and development of new drug targets for various disorders.

NOV14

A disclosed NOV14 nucleic acid of 986 nucleotides (also referred to as CG56023-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 14A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 23-25 and ending with a TGA codon at nucleotides 974-976. The start and stop codons are shown in bold in Table 14A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 14A. NOV14 nucleotide sequence (SEQ ID NO:59).

CTGGGGATTATGCCCATACTTATGGCTATAGGAACTGGACAGAAATAAGTGAATTTATCCTCATGAGCTT
CTCTTCCCTACCTACTGAAATACAGTCATTGCTCTTCCCTGACATTTCTAACTATCTATTTGGTTACTCTGAA
GGGAAACAGCCTCATCATTCTGGTTACCCTAGCTGACCCCATGCTACACAGCCCATGTACTTCTTCCCTCAG
AAACTTATCTTCTCTGGAGATTGGCTTCAACCTAGTCATTGTGCCAAAATGCTGGGGACCCGTCTTGCCCA
GGACACAACCATCTCCTTCCCTGGCTGTGCCACTCAGATGATTTCTTCTTCTTTGGGGTAGCTGAATG
CTTCTCCTGGCTACCATGGCATATGACCGCTATGTGGCCATCTGCAGTCCCTTGCACTACCCAGTCATCAT
GAACCAAAGGACACGGGCCAACTGGCTGCTGCTTCTGGTTCCAGGCTTCTCTGTAGCTACTGTGCAGAC
CACATGGCTCTTCAGTTTCCATTCTGTGGCACCACAAAGGTGAACCACTTCTGTGACAGCCCGCTGT
GCTGAAGCTGGTCTGTGCAGACACAGCACTGTTGAGATCTACGCCATCGTCGGAACCACTTCTGGTGGTCAT
GATCCCCTGCTTGTGATCTTGTGTCTCTATACTCGCATGTGCTGCTATCCTCAAGATCCCATCAGCTAA
AGGGAAGCATAAAGCCTTCTCTACGTGCTCCTCACACCTCCTTGTGTCTCTCTTTTCTATATATCTTCTAG
CCTCACCTACTTCTGGCCTAAATCAAATAATTCTCCTGAGAGCAAGAAGTTGTTATCATTATCCTACACTGT
TGTGACTCCCATGTTGAACCCATTATCTACAGCTTGAGAAATAGCGAGGTGAAGAATGCCCTCAGCAGGAC
CTTCCACAAGGTCCTAGCCCTCAGAACTGTATCCCATAGACCTTAGGAA

The disclosed NOV14 polypeptide (SEQ ID NO:60) encoded by SEQ ID NO:59 has 321 amino acid residues and is presented in Table 14B using the one-letter amino acid code.

Table 14B. Encoded NOV14 protein sequence (SEQ ID NO:60).

MPILMAIGNWTEISEFILMSFSSLPTEIQSLLFLTFLTYLVTLKGNLSLIILVTLADPMLHSPM
YFFLRNLSFLEIGFNLVIVPKMLGTLQAQTTISFLGCATQMYFFFFFGVAECFLLATMAYDRY
VAICSPHLHPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTNKNVHFFCDSPVVLKL
VCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAKGKHKAFSTCSSHLLVVSFLF
YISSSLTYFWPKSNNSPESKKLLSLSYTVVTPMLNPPIIYSLRNSEVKNALSRTFHKVLALRNCIP

A search of sequence databases reveals that the NOV14 amino acid sequence has 234 of 310 amino acid residues (75%) identical to, and 264 of 310 amino acid residues (85%) similar to, the 315 amino acid residue ptnr: SPTREMBL-ACC:Q9JKA6 protein from *Mus musculus* (Mouse) (OLFACTORY RECEPTOR P2) ($E = 4.0e^{-124}$).

The disclosed NOV14 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 14C.

Table 14C. BLAST results for NOV14

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14423805 sp Q9H207 OAA5_HUMAN	OLFACTORY RECEPTOR 10A5 (HP3)	317	317/317 (100%)	317/317 (100%)	e-154
gi 12007437 gb AAG45207.1 AF321237_4 (AF321237)	hP4 olfactory receptor [Homo sapiens]	317	300/317 (94%)	305/317 (95%)	e-145
gi 12007412 gb AAG45186.1 (AF321233)	P3 olfactory receptor [Mus musculus]	317	292/316 (92%)	302/316 (95%)	e-140
gi 15419583 gb AAK97076.1 AF293080_1 (AF293080)	olfactory receptor P3 [Mus musculus]	324	294/320 (91%)	304/320 (94%)	e-140
gi 12007411 gb AAG45185.1 (AF321233)	P4 olfactory receptor [Mus musculus]	317	281/316 (88%)	296/316 (92%)	e-136

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 14F. In the ClustalW alignment of the NOV14 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 14D. ClustalW Analysis of NOV14

- 1) Novel NOV14 (SEQ ID NO:60)
- 2) gi|14423805|sp|Q9H207|OAA5_HUMAN OLFACTORY RECEPTOR 10A5 (HP3) (SEQ ID NO:382)
- 3) gi|12007437|gb|AAG45207.1|AF321237_4 (AF321237) hP4 olfactory receptor [Homo sapiens] (SEQ ID NO:383)
- 4) gi|12007412|gb|AAG45186.1| (AF321233) P3 olfactory receptor [Mus musculus] (SEQ ID NO:384)
- 5) gi|15419583|gb|AAK97076.1|AF293080_1 (AF293080) olfactory receptor P3 [Mus musculus] (SEQ ID NO:385)
- 6) gi|12007411|gb|AAG45185.1| (AF321233) P4 olfactory receptor [Mus musculus] (SEQ ID NO:386)

			10	20	30	40	50	60	
NOV14	1	---	MPILMAIGNWTEISEFILMSFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	57				
gi 14423805	1	-----	MAIGNWTEISEFILMSFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	53				
gi 12007437	1	-----	IAIGNWTRISEFILMSFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	53				
gi 12007412	1	-----	MATGNOTRITTEFILMSFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	53				
gi 15419583	1	MLFMLIPMATGNOTRITTEFILMSFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	60					
gi 12007411	1	-----	MAAGNOTRIAEFILMGFSSLPTEIQSLLFLTFLTIYLVTLK	CNSLIILVTLAD	53				
			70	80	90	100	110	120	
NOV14	58	PMLHSPMYFFLRNLSFLEIGFNLVIVPKMLGTL	LAQDTTISFLGCATQMYFFFFFGVAEC	117					
gi 14423805	54	PMLHSPMYFFLRNLSFLEIGFNLVIVPKMLGTL	LAQDTTISFLGCATQMYFFFFFGVAEC	113					
gi 12007437	54	PMLHSPMYFFLRNLSFLEIGFNLVIVPKMLGTL	LAQDTTISFLGCATQMYFFFFFGVAEC	113					
gi 12007412	54	PMLQSPMYFFLRNLSFLEIGFNLVIVPKMLGTL	LAQDTTISFLGCATQMYFFFFFGVAEC	113					

gi 15419583	61	PMLQSPMYFFLRNLSFLEIGFNLVIVPKMLGTLTAQDTSISFLGCATQMYFFFFFGVAEC	120
gi 12007411	54	PMLQSPMYFFLRNLSFLEIGFNLVIVPKMLRRTLAQDTSISFLGCATQMYFFFFFGVAEC	113
5			
		130 140 150 160 170 180	
NOV14	118	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	177
gi 14423805	114	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	173
gi 12007437	114	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	173
gi 12007412	114	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	173
10			
gi 15419583	121	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	180
gi 12007411	114	FLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQTTWLFSFPFCGTN	173
		190 200 210 220 230 240	
15			
NOV14	178	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	237
gi 14423805	174	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	233
gi 12007437	174	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	233
gi 12007412	174	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	233
20			
gi 15419583	181	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	240
gi 12007411	174	KVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTRIAAAILKIPSAK	233
		250 260 270 280 290 300	
25			
NOV14	238	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	297
gi 14423805	234	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	293
gi 12007437	234	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	293
gi 12007412	234	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	293
30			
gi 15419583	241	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	300
gi 12007411	234	GKHKAFSTCSSHLLVVSFLFYISLTYFRPKSNNSPESKKLLSLSYTVVTPMLNPIIYSL	293
		310 320	
35			
NOV14	298	RNSEVKNALSRTHKVLALRNCTP	321
gi 14423805	294	RNSEVKNALSRTHKVLALRNCTP	317
gi 12007437	294	RNSEVKNALSRTHKVLALRNCTP	317
gi 12007412	294	RNSEVKNALSRTHKVLALRNCTP	317
40			
gi 15419583	301	RNSEVKNALSRTHKVLALRNCTP	324
gi 12007411	294	RNSEVKNALSRTHKVLALRNCTP	317

Table 14E lists the domain descriptions from DOMAIN analysis results against NOV14. This indicates that the NOV14 sequence has properties similar to those of other proteins known to contain this domain.

Table 14E Domain Analysis of NOV14

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 103 bits (256), Expect = 2e-23

45			
NOV14:	46	GNSLIILVTLADPMLHSPMYFFLRNLSFLEIGFNLVIVPKMLGTLTAQDTSISFLGCATQ	105
Sbjct:	1	GNLLVILVILRTTKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
50			
NOV14:	106	MYFFFFFGVAECFLLATMAYDRYVAICSPHYPVIMNQRTAKLAAASWFGFPVATVQT	165
Sbjct:	61	GALFVVNGYASILLTASIDRYLAIVHPLRYRRIPTPRAKVLILLVWVLAL-----	113
55			
NOV14:	166	TWLFSFPFCGTNKVNHFFCDSPPVLRLVCADTALFEIYAIVGTILVVMIPCLLILCSYTR	225
Sbjct:	114	--LLSLPPLLFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVPLLVILVCYTR	171

NOV14: 226 IA-----AAILKIPSAKGKHKAFSTCSSHLLVVSIFY----ISSSLTYFWPKSNNS 272
 | || + + ++ + + +
 Sbjct: 172 ILRTLKRKRARSQRSCLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSCLLSIWRV 231
 NOV14: 273 PESKKLLSLSYTVVTPMLNPITY 295
 + |++| | |||||
 Sbjct: 232 LPTALLITLWLAYVNSCLNPITY 254

10 G-Protein Coupled Receptor (GPCRs) have been identified as extremely large
 subfamily of G protein-coupled receptors in a number of species. These receptors share a
 seven transmembrane domain structure with many neurotransmitter and hormone receptors,
 and are likely to underlie the recognition and G-protein-mediated transduction of various
 signals. Previously, GPCR genes cloned in different species were from random locations in the
 15 respective genomes. The human GPCR genes are intron less and belong to four different gene
 subfamilies, displaying great sequence variability. These genes are dominantly expressed in
 olfactory epithelium.

 Olfactory receptors (ORs) have been identified as extremely large subfamily of G
 protein-coupled receptors in a number of species. These receptors share a seven
 20 transmembrane domain structure with many neurotransmitter and hormone receptors, and are
 likely to underlie the recognition and G-protein-mediated transduction of odorant signals.
 Previously, OR genes cloned in different species were from random locations in the respective
 genomes. The human OR genes are intron less and belong to four different gene subfamilies,
 displaying great sequence variability. These genes are dominantly expressed in olfactory
 25 epithelium.

 The disclosed NOV14 nucleic acid of the invention encoding a G-Protein Coupled
 Receptor -like protein includes the nucleic acid whose sequence is provided in Table 14A or a
 fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose
 bases may be changed from the corresponding base shown in Table 14A while still encoding a
 30 protein that maintains its G-Protein Coupled Receptor -like activities and physiological
 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids
 whose sequences are complementary to those just described, including nucleic acid fragments
 that are complementary to any of the nucleic acids just described. The invention additionally
 includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures
 35 include chemical modifications. Such modifications include, by way of nonlimiting example,
 modified bases, and nucleic acids whose sugar phosphate backbones are modified or
 derivatized. These modifications are carried out at least in part to enhance the chemical
 stability of the modified nucleic acid, such that they may be used, for example, as antisense
 binding nucleic acids in therapeutic applications in a subject.

The disclosed NOV14 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 14B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 14B while still encoding a protein that maintains its G-Protein
5 Coupled Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 12 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like
10 protein (NOV14) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV14 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV14 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases
20 (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial,
25 fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction,
30 ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV14 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV14 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV14 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV15

NOV15 includes three novel G-Protein Coupled Receptor -like proteins disclosed below. The disclosed sequences have been named NOV15a and NOV15b.

NOV15a

A disclosed NOV15a nucleic acid of 943 nucleotides (also referred to as CG56065-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 15A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 2-4 and ending with a TGA codon at nucleotides 935-937. The start and stop codons are shown in bold in Table 15A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 15A. NOV15a nucleotide sequence (SEQ ID NO:61).

AATGGCAGCAGAAAACCATTCTTTTGTGACTAAGTTTATTCTGGTTGGGCTAACAGAGAAGTCAG
AGCTACAGCTGCCCCCTCTTCCTCGTCTTCCTGGGAATCTATGTAGTCACAGTGCTGGGGAACTTG
GGCATGATCACACTGATTGGGCTCAGTTCTCACCTGCACACACCTATGTACTGTTTCCTCAGCAG
TCTGTCCTTCATTGACTTCTGCCATTCCACTGTCATTACCCCTAAGATGCTGGTGAACCTTTGTGA
CAGAGAAGAACATCATCTCCTACCCTGAATGCATGACTCAGCTCTACTTCTTCCTCGTTTTTGCT
ATTGCAGAGTGTCACATGTTGGCTGCAATGGCATATGACGGCTACGTGGCCATCTGTAGCCCCCTT
GCTGTACAGCATCATCATATCCAATAAGGCTTGCTTTTCTCTGATTTTAGTGGTGTATGTAATAG
GCCTGATTTGTGCGTCAGCTCATATAGGCTGTATGTTTAGGGTTCAATTCTGCAAATTTGATGTG
ATCAACCATTATTTCTGTGATCTTATTTCTATCTTGAAGCTCTCCTGTTCTAGTACTTACATTAA
TGAGTTACTGATTTTAATCTTTAGTGGAATTAACATCCTTGTCCCCAGCCTGACCATCCTCAGCT
CTTACATCTTCATCATTGCCAGCATCCTCCGCATTCGCTACACTGAGGGCAGGTCCAAAGCCTTC
AGCACTTGACAGCTCCACATCTCGGCTGTTTCTGTTTCTTTGGGTCTGCAGCATTTCATGTACCT
GCAGCCATCATCTGTGCTCAGCTCCATGGACCAGGGGAAAGTGTCTCTGTGTTTTATACTATTGTTG
TGCCCATGCTGAACCCCCTGATCTACAGCCTGAGGAATAAAGATGTCCACGTTGCCCTGAAGAAA
ACGCTAGGGAAAAGAACATTCTTATGAACAGAA

The disclosed NOV15a polypeptide (SEQ ID NO:62) encoded by SEQ ID NO:61 has 311 amino acid residues and is presented in Table 15B using the one-letter amino acid code.

Table 15B. Encoded NOV15a protein sequence (SEQ ID NO:62).

MAAENHSFVTKFILVGLTEKSELQLPLFLVFLGIYVVTVLGNLGMITLIGLSSHLHTPMYCFLS
SLSFIDFCHSTVITPKMLNVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDGYVAICS
PLLYSIIISNKACFSLILVVYVIGLICASAHI GCMFRVQFCKFDVINHYFCDLISILKLSCSST
YINELLILIFSGINILVPSLTILSSYIFIIASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAA
FMYLQPSVSSMDQGVSSVFYTI VVPMNLPLIYSLRNKDVHVALKKTTLGKRTFL

A search of sequence databases reveals that the NOV15a amino acid sequence has 235 of 311 amino acid residues (75%) identical to, and 270 of 311 amino acid residues (86%) similar to, the 311 amino acid residue ptrn: SPTREMBL-ACC:O35184 protein from *Rattus norvegicus* (Rat) (Olfactory Receptor) ($E = 9.9e^{-121}$).

NOV15b

A disclosed NOV15b nucleic acid of 943 nucleotides (also referred to as CG56065-02) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 15C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 2-4 and ending with a TGA codon at nucleotides 935-937. The start and stop codons are shown in bold in Table 15C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 15C. NOV15b nucleotide sequence (SEQ ID NO:63).

AA**TGGCAGCAGAAA**ACCATTCTTTTGTGACTAAGTTTATTCTGGTTGGGCTAACAGAGAAGTCAGAGCTACA
GCTGCCCCCTCTCCTCGTCTTCCTGGGAATCTATGTAGTCACAGTGCTGGGGAACCTGGGCATGATCACACT
GATTGGGCTCAGTTCTCACCTGCACACCTATGTACTGTTTCCTCAGCAGTCTGTCCTTCATTGACTTCTG
CCATTCCACTGTCATTACCCCTAAGATGCTGGTGAACCTTTGTGACAGAGAAGAACATCATCTCCTACCCTGA
ATGCATGACTCAGCTCTACTTCTTCCTCGTTTTTGTCTATTGCAGAGTGTACATGTTGGCTGCAATGGCATA
TGACGGCTACGTGGCCATCTGTAGCCCCGTGCTGTACAGCATCATCATATCCAATAAGGCTTCTTTTCTCT
GATTTTAGTGGTGTATGTAATAGGCCTGATTTGTGCGTCAGCTCATATAGGCTGTATGTTTAGGGTTCAATT
CTGCAAAATTTGATGTGATCAACCATTATTTCTGTGATCTTATTTCTATCTTGAAGCTCTCCTGTTCTAGTAC
TTACATTAATGAGTTACTGATTTTAATCTTTAGTGGAATTAACATCCTGTCCCCAGCCTGACCATCCTCAG
CTCTTACATCTTCATCATTGCCAGCATCCTCCGCATTGCTACACTGAGGGCAGGTCCAAAGCCTTCAGCAC
TTGCAGCTCCACATCTCGGCTGTTTCTGTTTTCTTTGGGTCTGCAGCATTGATGACCTGCAGCCATCATC
TGTCAGCTCCATGGACCAGGGGAAAGTGCTCTGTGTTTTATACTATTGTTGTGCCCGTGTGAACCCCT
GATCTACGCCTGAGGAATAAAGATGTCCACGTGCGCCTGAAGAAAACGCTAGGGAAAAGAACATTCTTATG
AACAGAA

In a search of public sequence databases, the NOV15b nucleic acid sequence, localized to chromosome 4, has 770 of 937 bases (82%) identical to a gb:GENBANK-ID:AF282271|acc:AF282271.1 mRNA from *Mus musculus* (odorant receptor K11 gene, complete cds) ($E = 5.2e^{-135}$).

The disclosed NOV15b polypeptide (SEQ ID NO:64) encoded by SEQ ID NO:63 has 311 amino acid residues and is presented in Table 15D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV15b has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV15b may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic

reticulum (membrane) with a certainty of 0.3000, or in the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV15b is between positions 41 and 42: VLG-NL.

Table 15D. Encoded NOV15b protein sequence (SEQ ID NO:64).

```
MAAENHSFVTKFILVGLTEKSELQLPLFLVFLGIYVVTVLGNLGMITLIGLSSHLHTPMYCFLLSSLSFIDFC
HSTVITPKMLVNFVTEKNIIISYPECMTQLYFFLVFAIAECHMLAAMAYDGYVAICSPVLYSIIISNKACFSL
ILVVYVIGLICASAHIGCMFRVQFCKFDVINHYFCDLISILKLSSTYINELLILFSGINILVPSLTILS
SYIFI IASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAAFMYLQPSVSSMDQGVSSVFYTIIVPVLNPL
IYSLRNKDVHVALKKTTLGKRTFL
```

A search of sequence databases reveals that the NOV15b amino acid sequence has 237 of 311 amino acid residues (76%) identical to, and 273 of 311 amino acid residues (87%) similar to, the 314 amino acid residue ptnr:TREMBLNEW-ACC:AAG39856 protein from *Mus musculus* (Mouse) (Odorant Receptor K11) ($E = 2.6e^{-125}$).

NOV15b is predicted to be expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV15a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 15E.

Table 15E. BLAST results for NOV15a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17472672 ref XP_061794.1 (XM_061794)	similar to odorant receptor K11 (H. sapiens) [Homo sapiens]	311	311/311 (100%)	311/311 (100%)	e-140

gi 11692519 gb AAG39856.1 AF282271_1 (AF282271)	odorant receptor K11 [<i>Mus musculus</i>]	314	239/311 (76%)	273/311 (86%)	e-110
gi 11692527 gb AAG39860.1 AF282275_1 (AF282275)	odorant receptor K15 [<i>Mus musculus</i>]	311	236/311 (75%)	271/311 (86%)	e-108
gi 17472662 ref XP_061790.1 (XM_061790)	similar to odorant receptor K4h11 (H. sapiens) [<i>Homo sapiens</i>]	593	233/301 (77%)	261/301 (86%)	e-105
gi 2317704 gb AAB66333.1 (AF010293)	olfactory receptor [<i>Rattus norvegicus</i>]	311	235/311 (75%)	270/311 (86%)	e-105

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 15F. In the ClustalW alignment of the NOV15 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 15F. ClustalW Analysis of NOV15

- 1) Novel NOV15a (SEQ ID NO:62)
- 2) Novel NOV15b (SEQ ID NO:64)
- 3) gi|17472672|ref|XP_061794.1| (XM_061794) similar to odorant receptor K11 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:387)
- 4) gi|11692519|gb|AAG39856.1|AF282271_1 (AF282271) odorant receptor K11 [*Mus musculus*] (SEQ ID NO:388)
- 5) gi|11692527|gb|AAG39860.1|AF282275_1 (AF282275) odorant receptor K15 [*Mus musculus*] (SEQ ID NO:389)
- 6) gi|17472662|ref|XP_061790.1| (XM_061790) similar to odorant receptor K4h11 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:390)
- 7) gi|2317704|gb|AAB66333.1| (AF010293) olfactory receptor [*Rattus norvegicus*] (SEQ ID NO:391)

				10	20	30	40	50	60	
									
NOV15a	1		-----							1
NOV15b	1		-----							1
gi 17472672	1		-----							1
gi 11692519	1		-----							1
gi 11692527	1		-----							1
gi 17472662	1		MVKGNHSTVTEFNLAGLTDKPELQLPLFLLFLGIYVVTVVGNLSMITLIGFSSHLHTPMY							60
gi 2317704	1		-----							1
				70	80	90	100	110	120	
									
NOV15a	1		-----							1
NOV15b	1		-----							1
gi 17472672	1		-----							1
gi 11692519	1		-----							1
gi 11692527	1		-----							1
gi 17472662	61		HFLSSLSFDILCQSSVITPKMLVNFVSEARNIISYPACMTQLYFFLVLVISECHMLAAMAY							120
gi 2317704	1		-----							1

			130	140	150	160	170	180	
5	NOV15a	1	1					
	NOV15b	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 17472672	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 11692519	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 11692527	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
10	gi 17472662	121	DHYIAICNPLLYHVAMSQVCSWMVVEVYFMGFIGATCSHSLHAKSAFLLTILSSYIFIV	180					
	gi 2317704	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
			190	200	210	220	230	240	
15	NOV15a	1	1					
	NOV15b	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 17472672	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 11692519	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 11692527	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
20	gi 17472662	181	ASILCIRSTEGRSKTFSTCSSHISAVSVFFGGTSRSRFQVLGLEVRVRLGGCPDAGQTP	240					
	gi 2317704	1	----- ----- ----- ----- ----- ----- ----- ----- -----	1					
			250	260	270	280	290	300	
25	NOV15a	1	17					
	NOV15b	1	----- ----- ----- ----- ----- ----- ----- ----- -----	17					
	gi 17472672	1	----- ----- ----- ----- ----- ----- ----- ----- -----	17					
	gi 11692519	1	----- ----- ----- ----- ----- ----- ----- ----- -----	20					
	gi 11692527	1	----- ----- ----- ----- ----- ----- ----- ----- -----	17					
30	gi 17472662	241	ETQPPVQSLFSGHRNLAPSARAMEKKNVQVPWTLAERMETVDKIVDPGNHSSVTESILAGL	300					
	gi 2317704	1	----- ----- ----- ----- ----- ----- ----- ----- -----	17					
			310	320	330	340	350	360	
35	NOV15a	18	TEKSELQLPLFLVFLGIYVVTVLGNLGMITLIGLSSHLHTPMYCFLSSLSFIDFCHSTVI	77					
	NOV15b	18	TEKSELQLPLFLVFLGIYVVTVLGNLGMITLIGLSSHLHTPMYCFLSSLSFIDFCHSTVI	77					
	gi 17472672	18	TEKSELQLPLFLVFLGIYVVTVLGNLGMITLIGLSSHLHTPMYCFLSSLSFIDFCHSTVI	77					
	gi 11692519	21	SEKPELQLPLFLFLFIGIYMITVAGNLGMITLIGLSSHLHTPMYYFLSSLSFIDFCHSTVI	80					
	gi 11692527	18	SKKPELQLPLFLFLFLGIYVVTVVGNLGMITLIGLSSHLHTPMYYFLSSLSFIDFCHSTVI	77					
40	gi 17472662	301	SEKPELQLPLFLFLFLGIYVVTVVGNLGMITLIGLSSHLHTPMYYFLSSLSFIDFCHSTVI	360					
	gi 2317704	18	SKKSELQMPFLFVFLGIYVVTVVGNLGMITLIGLSSLLHTPMYYFLSSLSFIDFCHSTVI	77					
			370	380	390	400	410	420	
45	NOV15a	78	TPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDGYVAICSPLLYSIIIS	137					
	NOV15b	78	TPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDGYVAICSPVLYSIIIS	137					
	gi 17472672	78	TPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDGYVAICSPLLYSIIIS	137					
	gi 11692519	81	TPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECYILAAMAYDRVVAICNPLLYNVTMS	140					
	gi 11692527	78	TPKMLVNFVIEKNIISYTCMAQLYFFLVFAIAECHMLAAMAYDRVVAICNPLLYNVTMS	137					
50	gi 17472662	361	TPKMLVNFATEKNIISYPECMAQLYFFLVFAIAECHMLAAMAYDCYVAICSPLLYNVIMS	420					
	gi 2317704	78	TPKMLVNFVAKNIISYTCMTQLYFFLVFAIAECHMLAAMAYDRVVAICNPLLYNAIMS	137					
			430	440	450	460	470	480	
55	NOV15a	138	NKACFSLILVVYVIGLICASAHIQCMFRVQFCRFDVINHYFCDLISILKLSCSSTYINEL	197					
	NOV15b	138	NKACFSLILVVYVIGLICASAHIQCMFRVQFCRFDVINHYFCDLISILKLSCSSTYINEL	197					
	gi 17472672	138	NKACFSLILVVYVIGLICASAHIQCMFRVQFCRFDVINHYFCDLISILKLSCSSTYINEL	197					
	gi 11692519	141	YQIYIFLISGVYIIGVICASAHTGFMVIRFCRFDVINHYFCDLIPPLKLAACSNYINEM	200					
	gi 11692527	138	YQIYISLIFGVYIIGVVCASAHTGFMIRIQFCNLEVINHYFCDLIPPLKLAHSSTYINEL	197					
60	gi 17472662	421	YHHCFWLTGVYILGIIIGSTIHTSFMLRFLFKTNVINHYFCDLIPPLKLSCSSTYINEL	480					
	gi 2317704	138	YQSYISLISGVYIIGVVCASAHTGFMIRSQFCNLDVINHYFCDLIPPLKLAHSSTYINEL	197					
			490	500	510	520	530	540	
65	NOV15a	198	LILIFSGINILVPSLTILSSYIFIIASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAAF	257					
	NOV15b	198	LILIFSGINILVPSLTILSSYIFIIASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAAF	257					
	gi 17472672	198	LILIFSGINILVPSLTILSSYIFIIASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAAF	257					
	gi 11692519	201	LILIFFGTNINIFVPILTILSSYIFIIASILRIRSTEGRSKAFSTCSSHILAVAVFFGSAAF	260					
	gi 11692527	198	LVLCEFTGNIVVPILTILSSYIFIIANILRIRSTEGRSKAFSTCSSHILAVAVFFGSAAF	257					
70	gi 17472662	481	LVLVLSAFNIILPALTILSSYIFIIASILRIHSTEGRSKAFSTCSSHILAVAVFFGSAAF	540					
	gi 2317704	198	VILICGTGNIIVPILTILSSYIFIIATILHIRSTEGRYKAFSTCSSHILAVAVFFGSAAF	257					

					550	560	570	580	590							
5	NOV15a	258	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	GKRT	FL	311
	NOV15b	258	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	GKRT	FL	311
	gi 17472672	258	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	GKRT	FL	311
	gi 11692519	261	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	ERKT	FM	314
	gi 11692527	258	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	ERKL	FM	311
10	gi 17472662	541	MYLQ	PSSV	SMDQ	RKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	HQTAC	-	593
	gi 2317704	258	MYLQ	PSSV	SMDQ	GKVSS	VFYTI	VVPM	LNPL	IYSL	RNKD	VHVA	LKKTL	ERKS	FL	311

Table 15G lists the domain description from DOMAIN analysis results against NOV15. This indicates that the NOV15 sequence has properties similar to those of other proteins known to contain this domain.

Table 15G Domain Analysis of NOV15

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
 CD-Length = 254 residues, 100.0% aligned
 Score = 86.7 bits (213), Expect = 2e-18

NOV15:	41	GNLGMITLIGLSSHLHTPMYCFLLSSLSFIDFCHSTVITPKMLVNFVTEKNIISYPECMTQ	100
Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
NOV15:	101	LYFFLVFAIAECHMLAAMAYDGYVAICSPLLYSIIISNKACFSLILVVYVIGLICASAHI	160
Sbjct:	61	GALFVVNGYASILLTALSIDRYLAIVHPLRYRRIPTRRAKVLILLVWVLALLLSLPPL	120
NOV15:	161	GCMFRVQFCKFDVINHYFCD----LISILKLSCSSTYINELLILIFSGINILVPSLTIL	215
Sbjct:	121	LFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPLLVLVLCYTRILRTLKRA	180
NOV15:	216	SSYIFIIASILRIRYTEGRSKAFSTCSSHISAVSVFFGSAAFMYL---QPSSVSSMDQG	271
Sbjct:	181	RSQ-----RSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSLCLLSIWRVLP	235
NOV15:	272	KVSSVFYTIIVVPMNLPIIY	290
Sbjct:	236	LLITLWLAYVNSCLNPIIY	254

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals.

5 Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV15 nucleic acid of the invention encoding a G-Protein Coupled
10 Receptor -like protein includes the nucleic acid whose sequence is provided in Table 15A, 15C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 15A or 15C while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
15 nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones
20 are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 18 percent of the bases may be so changed.

25 The disclosed NOV15 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 15B, or 15D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 15B, or 15D while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a
30 functional fragment thereof. In the mutant or variant protein, up to about 23 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV15) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV15 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV15 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV15 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV15 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV15 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in

understanding of pathology of the disease and development of new drug targets for various disorders.

NOV16a

A disclosed NOV16a nucleic acid of 891 nucleotides (also referred to as CG56067-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 16A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 5-7 and ending with a TAA codon at nucleotides 878-880. The start and stop codons are shown in bold in Table 16a, and the 5' and 3' untranslated regions, if any, are underlined.

Table 16A. NOV16a nucleotide sequence (SEQ ID NO:65).

GAAAATGTCAGCAGGAAACCATTCTCAGTGACTGAGTTCATTCTGGCTGGGCTCTCAGAACAGCCAGAGCT
CCAGCTGCGCCTCTTCTCCTGTTCTTAGGAATCTATGTGGTCACAGTGGTGGGCAACTTGAGCATGATCAC
ACTGATTGGGCTCAGTTCTCACCTGCATACCCCATGTACTATTTCTCAGTGGTCTGTCTTCATTGATAT
CTGCCATTCCACTATCATTACCCCAAATGCTGGTGAACCTTGTGACAGAGAAGAACATCATCTCCTACCC
TGAATGCATGACTCAGCTTTACTTCTTCTCATTCTTTGCTATTGCAGAGTGTACATGTTGGCTGTAACGGC
ATATGACCGCTATGTTGCCATCTGCAGCCCCCTGTGTTACAATGTCATCATGTCTATCACCAGTCTCTG
GCTCACAGTGGGAGTTTACATTTTGGCATCTTGGATCTACAATTCACACCGGCTTTATGTTGAGACTCTT
TTTGTGCAAGACTAATGTGATTAACCATTATTTTGTGATCTCTTCCCTCTCTTGGGGCTCTCTGCTCCAG
CACCTACATCAATGAATTACTGGTTCTGGTCTTGAGTGCATTTAACATCCTGACGCTGCCTTAACCATCT
TGCTTCTTACATCTTTATCATTGCCAGCATCTCCGCATTCGCTCCACTGAGGGCAGGTCCAAAGCCTTCAG
CACTTGCAGCTCCACATCTTGGCTGTTGCTGTTTCTTTGGGTCTGCAGCATTATGTACCTGCAGCCATC
ATCTGTGAGTCCATGGACAGGGGAAAGTGTCTGTGTTTATACTATTGTTGTGCCATGCTGAACCC
CCAATCTATAGCCTAAGAAATAAGGAT

In a search of public sequence databases, the NOV16a nucleic acid sequence, localized to chromosome 4, has 729 of 888 bases (82%) identical to a gb:GENBANK- ID:AF282293|acc:AF282293.1 mRNA from *Mus musculus* (odorant receptor K4h11 gene, complete cds) ($E = 9.8e^{-127}$).

The disclosed NOV16a polypeptide (SEQ ID NO:66) encoded by SEQ ID NO:65 has 311 amino acid residues and is presented in Table 16B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV16a has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV16A may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic reticulum (membrane) with a certainty of 0.3000, or in the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV16A is between positions 41 and 42: VVG-NL.

Table 16B. Encoded NOV16a protein sequence (SEQ ID NO:66).

MSAGNHSSVTEFILAGLSEQPELQLRLFLFLGLGIYVVTVVGNLSMITLIGLSSHLHTPMYYFLS

GLSFIDICHSTIITPKMLVNFVTEKNIISYPECMTQLYFFLIFAIAECHMLAVTAYDRYVAICS
 PLLYNVIMSYHHCFWLTGVYILGILGSTIHTGFMLRLFLCKTNVINHYFCDLFPLLGLSCSST
 YINELLVLVLSAFNLTALILASYIFIILASILRIRSTEGRSKAFSTCSSHILAVAVFFGSAA
 FMYLQPSSVSSMDQGVSSVFYTIIVFPLNPSIA

5 A search of sequence databases reveals that the NOV16a amino acid sequence has 232 of 287 amino acid residues (80%) identical to, and 253 of 287 amino acid residues (88%) similar to, the 307 amino acid residue ptnr:TREMBLNEW-ACC:AAG39878 protein from *Mus musculus* (Mouse) (Odorant Receptor K4H11) ($E = 5.1e^{-122}$).

NOV16a is predicted to be expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV16b

20 A disclosed NOV16b nucleic acid of 939 nucleotides (also referred to as CG56753-02) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 16C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TAG codon at nucleotides 934-936. The start and stop codons are shown in bold in Table 16C, and the 5' and 3' untranslated regions, if any, are underlined.

25

Table 16C. NOV16b nucleotide sequence (SEQ ID NO:67).

ATGTGCAGGAGAAAATAATTCCTCAGTGACTGAGTTTCATTCTGGCTGGGCTCTCAGAACGCCAGAGCTCCAG
 CTGCCCCCTTCCTCCTGTTCTTAGGAATCTATGTGGTCACAGTGGTGGGCAACCTGGGCATGACCACATCG
 ATTTGGCTCAGTCTCACCCTGCACCCCTATGTACTATTTCTCAGCAGTCTGCTCTCATTGACTCTCG
 CATTCCACTGTTTCATTACCCTAAGATGCTGGTGAACCTTTGTGCAGAGAGAACAATCATCTCCTACCTCGAA
 TGCATGACTCAGCTCTACTTCTTCCTCGTTTTTGCTATTGCAGAGTGTACATGTTGGCTGCAATGGCGTAT
 GACCCTTGCATGGCCATCTGTAGCCCTTGCTGCACAGTGTACATATCCAATAAGGCTGCTTTTCTCTG
 ATTTTAGGGTGTATATAATAAGGCCCTGGTTTTGTGCATCAGTTCATACAGACAGTATGTTTAGGGTTCAATTC
 TGCAAAATTTGATTTGATTAAACCATATTTCTGTGATCTTCTTCCCCCTCCTAAAGCTCTCTTGCTCTAGTATC
 TATGTCAACAAACTACTTATTCTATGTGTTGGTGCAATTAACATCCTTTGCTCCCGAGCCTGACCATCTTTGC
 TCTTACATCTTTATTATTGCCAGCATCTCCACATTCGCTCCACTGAGGGCAGGTCCAAGACCTTCAGCACT
 TGTAGCTCCCACATGTTGGCGGTTGTAATCTTTTTGGATCTGCAGCATTCATGTACTTGCAGCCATCTTCA
 ATCAGCTCCATGGACCAGGGGAAAGTATCCTCTGTGTTTTATACTATTATGTGCCCATGTTGAACCCCTG
 ATTTATAGCCTGAGGAATAAAGATGTCATGTTTCCCTGAAGAAAATGCTACAGAGAAGAACATTATGT**TA**

In a search of public sequence databases, the NOV16b nucleic acid sequence has 770 of 935 bases (82%) identical to a gb:GENBANK-ID:AF282271|acc:AF282271.1 mRNA from *Mus musculus* (odorant receptor K11 gene, complete cds) ($E = 1.3e^{-136}$).

5 The disclosed NOV16b polypeptide (SEQ ID NO:68) encoded by SEQ ID NO:67 has 311 amino acid residues and is presented in Table 16D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV16b has A signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV16b may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic reticulum (membrane) with a certainty of 0.3000, or in the endoplasmic reticulum (lumen) 10 with a certainty of 0.3000. The most likely cleavage site for NOV16b is between positions 41 and 42: VVG-NL.

Table 16D. Encoded NOV16b protein sequence (SEQ ID NO:68).

MSGENNSSVTEFILAGLSEQPELQPLPLFLFLGLIYVTVVGNLGMTTLIWLSSHLHTPMYFFLSSLSFIDFC
HSTVITPKMLVNFVTEKNIISYPECMTQLYFFLVFAIAECHMLAAMAYDRYMAICSPLLYSVIIISNKACFSL
ILGVYIIGLVCASVHTDSMFRVQFCKFDLINHYFCDLLPLLLKSCSSIYVKNLLILCVGAFNIVPSPSLTILC
SYIFIIASILHIRSTEGRSAFTSCSSHMLAVVIFGSAAFMYLQVPSISSMDQGVSSVFYTIIVPMLNPL
IYSLRNKDVHPSLLKKMLORRTLL

15 A search of sequence databases reveals that the NOV16b amino acid sequence has 238 of 311 amino acid residues (76%) identical to, and 274 of 311 amino acid residues (88%) similar to, the 314 amino acid residue ptrn:SPTREMBL-ACC:Q9EQB8 protein from *Mus musculus* (Mouse) (Odorant Receptor K11) ($E = 1.0e^{-127}$).

NOV16b is predicted to be expressed in at least the following tissues: Apical microvilli
20 of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma
cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral
tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary

artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV16a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 16E.

Table 16E. BLAST results for NOV16a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17472662 ref XP_061790.1 (XM_061790)	similar to odorant receptor K4h11 (H. sapiens) [<i>Homo sapiens</i>]	593	265/284 (93%)	267/284 (93%)	e-121
gi 11692519 gb AAG39856.1 AF282271_1 (AF282271)	odorant receptor K11 [<i>Mus musculus</i>]	314	223/287 (77%)	250/287 (86%)	e-104
gi 11692563 gb AAG39878.1 AF282293_1 (AF282293)	odorant receptor K4h11 [<i>Mus musculus</i>]	307	232/287 (80%)	253/287 (87%)	e-102
gi 17472672 ref XP_061794.1 (XM_061794)	similar to odorant receptor K11 (H. sapiens) [<i>Homo sapiens</i>]	311	226/287 (78%)	252/287 (87%)	e-102
gi 11692527 gb AAG39860.1 AF282275_1 (AF282275)	odorant receptor K15 [<i>Mus musculus</i>]	311	224/287 (78%)	246/287 (85%)	e-102

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 16F. In the ClustalW alignment of the NOV16 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

5
10
15
20
25
30
35
40
45
50
55
60
65

			310	320	330	340	350	360	
	NOV16a	18	SEQPELQLRLFLFLFLGIYVVTVVGNL	SMITLIGLSSHLHTPMYYFLSCLSFID	CHSTHI	77			
	NOV16b	18	SEQPELQLPLFLFLFLGIYVVTVVGNL	GMITLIGLSSHLHTPMYYFLSSLSFID	CHSTVI	77			
5	gi 17472662	301	SEQPELQLRLFLFLFLGIYVVTVVGNL	GMITLIGLSSHLHTPMYYFLSSLSFID	CHSTVI	360			
	gi 11692519	21	SEKPELQLPLFLFLFLGIYMITVAGNL	GMITLIGLSSHLHTPMYYFLSSLSFID	CHSTVI	80			
	gi 11692563	18	PTKPELQLPLFLFLFLGIYLVTVGNL	GMITLIGLSSHLHTPMYYFLSSLSFID	CHSTVI	77			
	gi 17472672	18	TEKSELQLPLFLFLFLGIYVVTVL	GNLGMITLIGLSSHLHTPMYCF	LSLSFIDCHSTVI	77			
10	gi 11692527	18	SKKPELOMLPLFLFLFLGIYVVTVVG	NLGMITLIGLSSHLHTPMYYFLSSLSFID	CHSTVI	77			
			370	380	390	400	410	420	
	NOV16a	78	TPKMLVNFVTEKNIISYPECMTQLYFF	LIFAI	AECHMLAVTAYDRYVAICSP	LLYNVIMS	137		
	NOV16b	78	TPKMLVNFVTEKNIISYPECMTQLYFF	LIFAI	AECHMLAAMAYDRYVAICSP	LLYSVIIIS	137		
15	gi 17472662	361	TPKMLVNFATEKNIISYPECMAQLYL	FSIFA	AECHMLAAMAYDCYVAICSP	LLYNVIMS	420		
	gi 11692519	81	TPKMLVNFVTEKNIISYPGCM	TQLYFFLIFAI	AECYILAAMAYDRYVAICN	PLLYNVTMS	140		
	gi 11692563	78	TPKMLVNFTEENIISYPECMTQLYFF	SLFAI	AECHMLAVMAYDRYVAICN	PLLYKVVMS	137		
	gi 17472672	78	TPKMLVNFVTEKNIISYPECMTQLYFF	LIFAI	AECHMLAAMAYDGYVAICSP	LLYSIIIS	137		
20	gi 11692527	78	TPKMLVNFVTEKNIISYTCMAQLYFF	LIFAI	AECHMLAAMAYDRYVAICN	PLLYNVTMS	137		
			430	440	450	460	470	480	
	NOV16a	138	YHHCFWLTVGVIYILGLIGSTIHTG	EMLRFL	CKTNVINHYFCDLFL	LLGLSCSSTYINEL	197		
	NOV16b	138	NKACFSLIIGVYIIGVICASVHTDS	MFVQFCKFDLINHYFCDLFL	LLKLSCSSTYVVKL	197			
25	gi 17472662	421	YHHCFWLTVGVIYILGLIGSTIHTG	EMLRFL	CKTNVINHYFCDLFL	LLGLSCSSTYINEL	480		
	gi 11692519	141	YQIYIFLISGVYIIGVICASAH	TGFMVRIR	CKLDVINHYFCDLFL	LLKLACSN	TYINEM	200	
	gi 11692563	138	HHVCFWFTVGVIYILGLIGSSVHTG	LMLKLF	CKTNKINHYFCDLFL	LLGLSCSSTYINEL	197		
	gi 17472672	138	NKACFSLILVVYVIGLICASAHTG	CMFRVQFCKFDVINHYFCDL	ISLKLSCSSTYINEL	197			
30	gi 11692527	138	YQIYTSLIIGVYIIGVICASAHTG	FMIRIOF	CNLEVINHYFCDLFL	LLLELAHSSTYVVKL	197		
			490	500	510	520	530	540	
	NOV16a	198	LVLVLSAFNIIIPALTILASYIFI	IIASILRIR	STEGRSKAFSTCSSHILAVAV	FFGSAAF	257		
	NOV16b	198	LILCVGAFNIIIPSLTILCSYIFI	IIASILHIRSTEGRSKAFSTCSSH	MLAVVIFFGSAAF	257			
35	gi 17472662	481	LVLVLSAFNIIIPALTILASYIFI	IIASILRIHIRSTEGRSKAFSTCSSH	ILAVAVFFGSAAF	540			
	gi 11692519	201	LILFFGTINIFVPIITITISYIFI	IIASILRIRSTEGRSKAFSTCSSH	ILAVAVFFGSLAF	260			
	gi 11692563	198	LVLFLSALNIIIPALTILMSYILI	IIISILRIRSTEGRSKAFSTCSSH	ISAVAFVFGSAAF	257			
	gi 17472672	198	LILIFSGINIIIPSLTILSSYIFI	IIASILRIRYTEGRSKAFSTCSSH	ISAVSVFFGSAAF	257			
40	gi 11692527	198	LVLFCGTENIVVPTMITLISYIFI	IIANILRIRSTGCRSKAFSTCSSH	ILAVAVFFGSAAF	257			
			550	560	570	580	590		
	NOV16a	258	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	291	
	NOV16b	258	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	311	
45	gi 17472662	541	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	593	
	gi 11692519	261	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	314	
	gi 11692563	258	TYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	307	
	gi 17472672	258	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	311	
50	gi 11692527	258	MYLQPVSSVSSMDQGVSSVFYTI	IVVPM	LNPLIYSLRNKDVHVS	KKMLQRR	TL	311	

Table 16G lists the domain description from DOMAIN analysis results against NOV16. This indicates that the NOV16 sequence has properties similar to those of other proteins known to contain this domain.

protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally
5 includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense
10 binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 18 percent of the bases may be so changed.

The disclosed NOV16 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 16B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding
15 residue shown in Table 16B while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 23 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

20 The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV16) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV16 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
25 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV16 nucleic acids and proteins of the invention are useful in potential
30 therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to

starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy(DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV16 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV16 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV16 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV17

NOV17 includes three novel G-Protein Coupled Receptor -like proteins disclosed below. The disclosed sequences have been named NOV17a, NOV17b, NOV17c, and NOV17d.

NOV17a

A disclosed NOV17a nucleic acid of 962 nucleotides (also referred to as CG56657-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 17A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 18-20 and ending with a TAG codon at nucleotides 954-956. The start and stop codons are shown in bold in Table 17A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 17A. NOV17a nucleotide sequence (SEQ ID NO:69).

GATCGTATGAATGCCCCATGGAAAAATACAATCAAACGTCAACTGATTTCATCTTATTGGGGCTGTTCCAC
CATCAAAAATTGGCCTTTTCCTCTTCATTCTCTTGTTCATTTTCCTAATGGCTCTAATTGGAAACCTAT
CCATGATTCTTCTCATCTTCTTGGACACCCATCTCCACACACCCATGTATTTCCCTGCTTAGTCAGCTCTCCC
TCATTGACCTAAATACATCTCTACGATTGTTCTTAAGATGGCTTCTGATTTTCTGTATGGAAACAAGTCTA
TCTCCTTCATTGGGTGTGGGATTGAGAGTTTCTTCTTCATGACTTTTGCAGGTGCAGAAGCGCTGCTCCTGA
CATCAATGGCCTATGATCGTTATGTGGCCATTTGCTTTTCTCCTCTCCACTATCCCATCCGTATGAGCAAAAGAA
TGTATGTGCTGATGATAACAGGATCTTGGATGATAGGCTCCATCAACTCTTGTGCTCACACAGTATATGCAT
TCCGTATCCCATATTGCAAGTCCAGAGCCATCAATCATTTTCTGTGATGTTCCAGCTATGTTGACATTAG
CCTGTACAGACACCTGGGTCTATGAGTACACAGTGTTTTGTAGCAGCACCATCTTCTTGTGTTTCCCTTCA
CTGGCATTGCGTGTTCCTATGGCTGGGTCTCCTTGTGCTTACCGCATGCACCTTGCAGAAGGGAGGAAAA
AGGCTATTTCGACCTGCAGCACCCACCTCACGTAGTAACCTTCTACTATGCACCTTTGCTTATACCTATC
TATGTCCAAGATCCCTGCGATCTCTGACAGAGGACAAGGTCTGGCTGTTTCTACACCATCCTCACCCCAA
TGCTCAACCCCATCATCTACAGCCTGAGAAACAAGGAGGTGATGGGGGCCCTGACACGAGTGATTGAGAATA
TCTTCTCGGTGAAAAATGTAGACATAC

The disclosed NOV17a polypeptide (SEQ ID NO:70) encoded by SEQ ID NO:69 has 312 amino acid residues and is presented in Table 17B using the one-letter amino acid code.

Table 17B. Encoded NOV17a protein sequence (SEQ ID NO:70).

MENYNQTSTDFILLGLFPFSKIGLFLFILFVLIFLMALIGNLSMILLIFLDTHLHTPMYFLLSQ
LSLIDLNYISTIVPKMASDFLYGNKISIFIGCGIQSFFFMTFAGAEALLTSMAYDRYVAICFP
LHYPIRMSKRMVYVLMITGSWMIGSINCAHTVYAFRIPYCKSRRAINHHFFCDVPAMLTACTDTW
VYEYTVFLSSTIFLVFPFTGIACSYGWVLLAVYRMHSAEGRKKAYSTCSTHLTVVTFYYPAY
TYLCPRSLRSLTEDKVLAVFYTILTPMLNPIIYSLRNKEVMGALTRVIQNIQSVKM

A search of sequence databases reveals that the NOV17a amino acid sequence has 148 of 305 amino acid residues (48%) identical to, and 192 of 305 amino acid residues (62%) similar to, the 316 amino acid residue ptrn: TREMBLNEW-ACC:AAG45196 protein from *Mus musculus* (Mouse) (T2 OLFACTORY RECEPTOR) ($E = 8.0e^{-73}$).

NOV17b

A disclosed NOV17b nucleic acid of 962 nucleotides (also referred to as CG56657-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 17C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 18-20 and ending with a TAG codon at nucleotides 954-956. The start and stop codons are shown in bold in Table 17C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 17C. NOV17b nucleotide sequence (SEQ ID NO:71).

GATCGTATGAATGCCCCATGAAAAATTACAATCAAACGTCAACTGATTTCATCTTATTGGGGCTGTTCCAC
CATCAAAAATTGGCCTTTTCCTCTTCATTCTCTTTGTTCTCATTTTCTAATGGCTCTAATTGGAAACCTAT
CCATGATTCTTCTCATCTTCTTGGACACCCATCTCCACACACCCATGTATTTCTGCTTAGTCAGCTCTCCC
TCATTGACCTAAATTACATCTCTACGATTGTTCTTAAGATGGCTTCTGATTTTCTGTATGGAAACAAGTCTA
TCTCCTTCATTGGGTGTGGGATTGAGAGTTTCTTCTTCATGACTTTTGCAGGTGAGAAAGCGCTGCTCCTGA
CATCAATGGCCTATGATCGTTATGTGGCCATTTGCTTTCTCTCCGCTATCCCATCCGTATGAGCAAAAGAA
TGTATGTGCTGATGATAACAGGATCTTGGATGATAGGCTCCATCAACTCTTGTGCTCACACAGTATATGCAT
TCCGTATCCCATATTGCAAGTCCAGAGCCATCAATCATTTTCTGTGATGTTCCAGCTATGTTGACATTAG
CCTGTACAGACACCTGGGTCTATGAGTACACAGTGTTTTGGAGCAGCACCATCTTCTTGTGTTTCCCTTCA
CTGGCATTGCGTGTTCCTATGGCTGGGTTCCTCTGCTGTCTACCGCATGCACCTGTCAGAAAGGGAGGAAAA
AGGCCTATTGACCTGCAGCACCACCTCACTGTAGTAACCTTTCTACTATGCACCCTTTGCTTATACCTATC
TATGTCCAAGATCCCTGCGATCTCTGACAGAGGACAAGGTCTGGCTGTTTTCTACACCATCTCACCCCAA
TGCTCAACCCCATCATCTACAGCCTGAGAAACAAGGAGGTGATGGGGGCCCTGACACGAGTGATTAGAATA
TCTTCTCGGTGAAAATGTAGACATAC

In a search of public sequence databases, the NOV17b nucleic acid sequence, localized to chromosome 4, has 321 of 342 bases (93%) identical to a gb:GENBANK-ID:HSHTPRH07|acc:X64978.1 mRNA from *Homo sapiens* (*H.sapiens* mRNA HTPCRH07 for olfactory receptor) ($E = 2.9e^{-62}$).

The disclosed NOV17b polypeptide (SEQ ID NO:72) encoded by SEQ ID NO:71 has 311 amino acid residues and is presented in Table 17D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV17b has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.4600. Alternatively, NOV17b may also localize to the microbody (peroxisome) with a certainty of 0.2311, the endoplasmic reticulum (membrane) with a certainty of 0.1000, or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV17B is between positions 43 and 44: NLS-MI.

Table 17D. Encoded NOV17b protein sequence (SEQ ID NO:72).

MENYNQTSTDFILLGLFPSPKIGLFLFILFVLIFLMALIGNLSMILLIFLDTHLHTPMYFLLSQLSLIDLNY
ISTIVPKMASDFLYGNKSIISFIGCGIQSFFMTFAGAEALLTSMAYDRYVAICFPLRYPIRMSKRMVLM
TGSWMIGSINSCHAVYAFRIPIYCKSRRAINHFCDVPAMLTACTDTWVVEYTVFLSSTIFLVFPFTGIACS
YGVVLLAVYRMHSAEGRKKAYSTCSTHLTVVTFFYAPFAYTYLCPRSLRSLTEDKVLAVFYITILPMLNP
IYSLRNKEVMGALTRVIONIFSVM

A search of sequence databases reveals that the NOV17b amino acid sequence has 148 of 305 amino acid residues (48%) identical to, and 191 of 305 amino acid residues (62%) similar to, the 316 amino acid residue ptnr:TREMBLNEW-ACC:AAG45196 protein from *Mus musculus* (Mouse) (T2 Olfactory Receptor) ($E = 8.0e^{-73}$).

NOV17b is predicted to be expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral

VAEGLLLTSMAYDRYVAICFPLHYPIRISKRVCMITGSSWISSINCAHTVYALCIPYCKSRAINHFFCD
VPAMLTACTDTWVYESTVFLSSTIFLVLPFTGIACSYGRVLLAVYRMHSAEGRKKAYSTCSTHLTVVSFYF
APFAYTYVRPRSLRSPTEDEKILAVFYTILTPMLNPITYSLRNKEVMGVLTVQVIQKIFSVKM

A search of sequence databases reveals that the NOV17d amino acid sequence has 138 of 269 amino acid residues (51%) identical to, and 183 of 269 amino acid residues (68%) similar to, the 316 amino acid residue ptnr:SPTREMBL-ACC:Q9D3U9 protein from *Mus musculus* (Mouse) (4933433E02rik Protein) ($E = 3.9e^{-71}$).

NOV17d is predicted to be expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV17a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 17I.

Table 17I. BLAST results for NOV17a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
>gi 17445356 ref XP_060561.1 (XM_060561)	similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [<i>Homo sapiens</i>]	312	312/312 (100%)	312/312 (100%)	e-149
gi 17445348 ref XP_060559.1 (XM_060559)	similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [<i>Homo sapiens</i>]	533	199/233 (85%)	206/233 (88%)	1e-95

NOV17d 277 ----- 277
gi|17445356| 312 ----- 312
gi|17445348| 533 ----- 533
gi|17437047| 472 ----- 472
5 gi|17437056| 695 ----- 695
gi|17456595| 549 SIRRQREFMPEEKDVTYWEKRRKNNEAAKRSREKRRRLNDAAIEGRLAALMEENALLKGE 608

1150 1160 1170
.....|.....|.....|.....|.....|.....|
10 NOV17a 312 ----- 312
NOV17b 312 ----- 312
NOV17c 277 ----- 277
NOV17d 277 ----- 277
15 gi|17445356| 312 ----- 312
gi|17445348| 533 ----- 533
gi|17437047| 472 ----- 472
gi|17437056| 695 ----- 695
gi|17456595| 609 LKALKLRFGLLPLTGSAGSPLDWGPPAWG 638

Table 17F lists the domain description from DOMAIN analysis results against NOV17. This indicates that the NOV17 sequence has properties similar to those of other proteins known to contain this domain.

Table 17F Domain Analysis of NOV17

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 99.4 bits (246), Expect = 3e-22

NOV17: 40 GNLSMILLIFLDTHLHTPMYFLLSQLSLIDLNYISTIVPKMASDFLYGNKSISFIGCGIQ 99
||| +||+| || | ++ || ++ |+ | + | +
Sbjct: 1 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV 60
NOV17: 100 SFFMTFAGAEALLTSMAYDRYVAICFPLHYPIRMSKRMVVLMTGSMIGSINCAHT 159
|+ | |||+++ |||+|| || | + | ++| ||+ + |
Sbjct: 61 GALFVNGYASILLTASIDRYLAIVHPLRYRRIPTPRRAKVLILLVWVLALLLSLPPL 120
NOV17: 160 VYAFRIPYCKSRAINHFFCDVPAMLTACTDTWVYEYTVFLSSTIFLVFPFTGIACSYGW 219
++ + + + + | | | + + | | | |
Sbjct: 121 LF-----SWLRTVEEGNTTVCLIDFPBESVKRSYVLLSTLVGFVLPPLLVILVCYTR 171
NOV17: 220 VLLAV-----YRMHSAEGRKKAYSTCSTHLLTVVTFYY---APFAYTYLCPRSLRS 266
+| + + + || | + | + + |
Sbjct: 172 ILRTLKRKRSQRSLSKRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSLCLLSIWRV 231
NOV17: 267 LTEDKVLAVFYTILTPMLNPIIY 289
| ++ ++ + |||||
Sbjct: 232 LPTALLITLWLAYVNSCLNPIIY 254

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the

respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large subfamily of G
5 protein-coupled receptors in a number of species. These receptors share a seven
transmembrane domain structure with many neurotransmitter and hormone receptors, and are
likely to underlie the recognition and G-protein-mediated transduction of odorant signals.
Previously, OR genes cloned in different species were from random locations in the respective
genomes. The human OR genes are intron less and belong to four different gene subfamilies,
10 displaying great sequence variability. These genes are dominantly expressed in olfactory
epithelium.

The disclosed NOV17 nucleic acid of the invention encoding a G-Protein Coupled
Receptor -like protein includes the nucleic acid whose sequence is provided in Table 17A,
17C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of
15 whose bases may be changed from the corresponding base shown in Table 17A or 17C while
still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and
physiological functions, or a fragment of such a nucleic acid. The invention further includes
nucleic acids whose sequences are complementary to those just described, including nucleic
acid fragments that are complementary to any of the nucleic acids just described. The
20 invention additionally includes nucleic acids or nucleic acid fragments, or complements
thereto, whose structures include chemical modifications. Such modifications include, by way
of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones
are modified or derivatized. These modifications are carried out at least in part to enhance the
chemical stability of the modified nucleic acid, such that they may be used, for example, as
25 antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or
variant nucleic acids, and their complements, up to about 7 percent of the bases may be so
changed.

The disclosed NOV17 protein of the invention includes the G-Protein Coupled
Receptor -like protein whose sequence is provided in Table 17B or 17D. The invention also
30 includes a mutant or variant protein any of whose residues may be changed from the
corresponding residue shown in Table 17B or 17D while still encoding a protein that
maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a
functional fragment thereof. In the mutant or variant protein, up to about 54 percent of the
residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV17) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV17 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV17 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV17 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV17 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV17 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in

assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV18

5 NOV18 includes three novel G-Protein Coupled Receptor -like proteins disclosed below. The disclosed sequences have been named NOV18a and NOV18b.

NOV18a

A disclosed NOV18a nucleic acid of 1062 nucleotides (also referred to as CG56663-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 18A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 10-12 and ending with a TAA codon at nucleotides 948-950. The start and stop codons are shown in bold in Table 18A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 18A. NOV18a nucleotide sequence (SEQ ID NO:77).

TAGAGATGGATGGAACCAATGGCAGCACCCAAACCCATTTTCATCCTACTGGGATTCTCTGACCGACCCCATC
TGGAGAGGATCCTCTTTGTGGTCATCCTGATCGGTACCTCCTGACCCCTCGTAGGCAACACCACCATCATCC
TGGTGTCCCGGCTGGACCCACCTCCACACCCCATGTAATCTTCTCGCCACCTTTCTCTTCTGGACC
TCAGTTTCACCAACAGCTCCATCCCCAGCTGCTCTACAACCTTAATGGATGTGACAAGACCATCAGCTACA
TGGGCTGTGCCATCCAGCTCTTCTGTTCCTGGGTCTGGGTGGTGTGGAGTGCCTGCTTCTGGCTGTCTATGG
CCTATGACCGGTGTGTGGCTATCTGCAAGCCCTGCACTACATGGTGATCATGAACCCAGGCTCTGCCGGG
GCTTGGTGTCTAGTGACCTGGGGCTGTGGGTGGCCAACTCCTTGGCCATGTCTCTGTGACCTGCGCTTAC
CCCGCTGTGGGCACCACGAGGTGGACCACTTCTGCGTGAGATGCCGCCCTGATCCGGATGGCCTGCGTCA
GCACTGTGGCCATCGAAGGCACCGTCTTTGTCTGAAAAAGGTGTGTGTCTGTCCCCCTTGGTGTTTATCC
TGCTCTCTTACAGCTACATTTGTGAGGGCTGTGTACAAATTCGGTCAGCATCAGGAAGGCAGAAGGCCTTCG
GCACCTGCGGCTCCCATCTCACTGTGGTCTCCCTTTTCTATGGAACATCATCTACATGTACATGCAGCCAG
GAGCCAGTTCTTCCAGGACCAAGGCATGTTCTCTCATGCTCTTCTACAACATTGTCACCCCTCTCTCAATC
CTCTCATCTACCCCTCAGAAACAGAGAGGTGAAGGGGGCACTGGGAAGGTTGCTTCTGGGGAAGAGAGAGC
TAGGAAAGGAGTAAAGGCATCTCCACCTGACTTCACTTCCATCCAGGGCCACTGGCAGCATCTGGAACGGCT
GAATTCAGCTGATATTAGCCACGACTCCCACTTGCCTTTTCTGGACTTTT

15 The disclosed NOV18a polypeptide (SEQ ID NO:78) encoded by SEQ ID NO:77 has 314 amino acid residues and is presented in Table 18B using the one-letter amino acid code.

Table 18B. Encoded NOV18a protein sequence (SEQ ID NO:78).

MDGTNGSTQTHFILLGFSDRPHLERILFVVILIAYLTLVGNTTIIILVSRLDPLHPTMPYFFLA
HLSFLDLSFTTSSIPLLNLNGCDKTI SYMGCAIQFLFLGLGGVECLLLAVMAYDRCVAICK
PLHYMVIMNPRLCRGLVSVTWGCGVANSLAMSPVTLRLPRCGHHEVDHFLREMPALIRMACVST
VAIEGTVFVLKKGVLSPLVFILLYSYIVRAVLQIRSASGRQKAFGTGSHLTVVSLFYGNII
YMXMQPGASSQDQGMFLMLFYNIIVTPLLNPLIYTLRNREVKGALGRLLLGKRELKGE

20 A search of sequence databases reveals that the NOV18a amino acid sequence has 194 of 237 amino acid residues (81%) identical to, and 215 of 237 amino acid residues (90%)

similar to, the 237 amino acid residue ptnr: SPTREMBL-ACC:Q9R0G5 protein from *Marmota marmota* (European marmot) (Olfactory Receptor) ($E = 3.5e^{-102}$).

NOV18b

A disclosed NOV18b nucleic acid of 1062 nucleotides (also referred to as CG56663-02) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 18C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 6-8 and ending with a TAA codon at nucleotides 948-950. The start and stop codons are shown in bold in Table 18C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 18C. NOV18b nucleotide sequence (SEQ ID NO:79).

TAGAGATGGATGGAACCAATGGCAGCACCCAAACCCATTTTCATCCTACTGGGATTCTTGACCGACCCCATC
TGGAGAGGATCCTCTTTGTGGTCATCCTGATCGCGTACCTCCTGACCCCTCGTAGGCAACACCACCATCATCC
TGGTGTCCCGGCTGGACCCCCACCTCCACACCCCATGTACTTCTTCCTCGCCACCTTTCTTCTTGGACC
TCAGTTTCACCACCAGCTCCATCCCCCAGCTGCTCTACAACCTTAATGGATGTGACAAGACCATCAGCTACA
TGGGCTGTGCCATCCAGCTCTTCTGTTCCTGGGTCTGGGTGGTGTGGAGTGCCTGCTTCTGGCTGTGATGG
CCTATGACCGGTGTGTGGCTATCTGCAAGCCCTGCATACATGGTGATCATGAACCCAGGCTCTGCCGGG
GCTTGGTGTGAGTGACCTGGGGCTGTGGGGTGGCCAACTCCTTGGCCATGTCTCCTGTGACCCCTGCGCTTAC
CCCCGTGTGGGCACCAAGAGGTGGACCACTTCTGCGTGAGATGCCCGCCCTGATCCGGATGGCCTGCGTCA
GCACTGTGGCCATCGACGGCACCGTCTTTGTCTTGGCGGTGGGTGTGTGTGCTGTCCCCCTTGGTGTATTATCC
TGCTCTCTTACAGCTACATTGTGAGGGCTGTGTTACAAATTCCGTCAGCATCAGGAAGGCAGAAGGCCTTCG
GCACCTGCGGCTCCCATCTCACTGTGGTCTCCCTTTTCTATGAAACATCATCTACATGTACATGCAGCCAG
GAGCCAGTTCTTCCAGGACCAGGGCATGTTCTCATGCTCTTCTACAACATTGTCAACCCCTCCTCAATC
CTCTCATCTACACCTCAGAAACAGAGAGGTGAAGGGGGCACTGGGAAGGTTGCTTTTGGGGAAGAGAGAGC
TAGGAAAGGAGTAAAGGCATCTCCACCTGACTTCACTTCCATCCAGGGCCACTGGCAGCATCTGGAACGGCT
GAATTCCAGCTGATATTAGCCCACGACTCCCAACTTGCCCTTTTCTGGACTTTT

In a search of public sequence databases, the NOV18b nucleic acid sequence has 600 of 710 bases (84%) identical to a gb:GENBANK-ID:AX008326|acc:AX008326.1 mRNA from *Marmota marmota* (Sequence 24 from Patent WO9967282) ($E = 8.8e^{-109}$).

The disclosed NOV18D polypeptide (SEQ ID NO:80) encoded by SEQ ID NO:79 has 314 amino acid residues and is presented in Table 18D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV18b has A signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV18b may also localize to the Golgi body with a certainty of 0.4000, the endoplasmic reticulum (membrane) with a certainty of 0.3000, or in the endoplasmic reticulum (lumen) with a certainty of 0.3000. The most likely cleavage site for NOV18b is between positions 42 and 43: LVG-NT.

Table 18D. Encoded NOV18b protein sequence (SEQ ID NO:80).

MDGTNGSTQTHFILLGFSDRPHLERILFVVILIAVLLTLVGNTTIIIVSRLDPHLHTPMYFFLAHLSFLDLS

FTTSSIPQLLYNLNGCDKTI SYMGCAIQFLFLGLGGVECLLLAVMAYDRCVAICKPLHYMVIMNPRLCRGL
VSVTWGCGVANSLAMSPVTLRLPRCGHHEVDHFLREMPALIRMACVSTVAIDGTVFVLAVGVVLSPLVFILL
SYSYIVRAVLQIRSASGRQKAFGTGSHLTVVSLFYGNIIYMYMQPGASSQDQGMFLMLFYNIIVTPLENPL
IYTLRNREVKGALGRLLLGKRELKGE

A search of sequence databases reveals that the NOV18b amino acid sequence has 183 of 305 amino acid residues (60%) identical to, and 237 of 305 amino acid residues (77%) similar to, the 320 amino acid residue ptnr:SPTREMBL-ACC:Q9Y3N9 protein from *Homo sapiens* (Human) (DJ88J8.1 (Novel 7 Transmembrane Receptor (Rhodopsin Family) (Olfactory Receptor Like) Protein) (HS6M1-15))) ($E = 2.8e^{-98}$).

NOV18b is predicted to be expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV18a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 18E.

Table 18E. BLAST results for NOV18a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17445344 ref XP_060558.1 (XM_060558)	similar to olfactory receptor (H. sapiens) [<i>Homo sapiens</i>]	314	314/314 (100%)	314/314 (100%)	e-164
gi 5901478 gb AAD55304.1 AF044033_1 (AF044033)	olfactory receptor [<i>Marmota marmota</i>]	237	194/237 (81%)	215/237 (89%)	2e-99
gi 13624329 ref NP_112165.1 (NM_030903)	olfactory receptor, family 2, subfamily W, member 1 [<i>Homo sapiens</i>]	320	184/305 (60%)	236/305 (77%)	1e-94

	gi 12054429	121	DRFTAICKPLHYFVVMNPHLCLKMIIMISISLANSVVLCTLTLNLPCTCGNNILDHFLCE	180			
		190	200	210	220	230	240
5	NOV18a	181	MPALIRMACVSTVAIEGTVFVLKKGVLSPLVFILLSSYSYIVRAVLQIRASGRQKAFGT	240			
	NOV18b	181	MPALIRMACVSTVAIDGTVFVLAVGVLSPLVFILLSSYSYIVRAVLQIRASGRQKAFGT	240			
	gi 17445344	181	MPALIRMACVSTVAIEGTVFVLKKGVLSPLVFILLSSYSYIVRAVLQIRASGRQKAFGT	240			
	gi 5901478	124	MPALIRMACVNTVAIEGTVFVLAVGIVLSPLVFILLSSYGHIVRAVFRIOSSSGRHRI FNT	183			
	gi 13624329	181	LPALVKIACVDTTTVEMSVFALGIITVLTPLILLILISYGYIAKAVLRTKSKASORKAMNT	240			
10	gi 12054431	181	LPALVKIACVDTTTVEMSVFALGIITVLTPLILLILISYGYIAKAVLRTKSKASORKAMNT	240			
	gi 12054429	181	LPALVKIACVDTTTVEMSVFALGIITVLTPLILLILISYGYIAKAVLRTKSKASORKAMNT	240			
		250	260	270	280	290	300
15	NOV18a	241	CGSHLTVVSLFYGNIIYMYMQPGASSSQDQCMFLMLFYNIIVTPLLNLPIYTLRNREVKGA	300			
	NOV18b	241	CGSHLTVVSLFYGNIIYMYMQPGASSSQDQCMFLMLFYNIIVTPLLNLPIYTLRNREVKGA	300			
	gi 17445344	241	CGSHLTVVSLFYGNIIYMYMQPGASSSQDQCMFLMLFYNIIVTPLLNLPIYTLRNREVKGA	300			
	gi 5901478	184	CGSHLTVVSLFYGNIIYMYMQPGSRSSQDQCKFLTFLFYNIIVTPLLNLPIYTLRNREVKGA	237			
	gi 13624329	241	CGSHLTVVSMFYGTIIYMYLQPCNRASKDQCKFLTFLFYTVITPSLNLPIYTLRNKDMKDA	300			
20	gi 12054431	241	CGSHLTVVSMFYGTIIYMYLQPCNRASKDQCKFLTFLFYTVITPSLNLPIYTLRNKDMKDA	300			
	gi 12054429	241	CGSHLTVVSMFYGTIIYMYLQPCNRASKDQCKFLTFLFYTVITPSLNLPIYTLRNKDMKDA	300			
		310	320				
25	NOV18a	301	LGRLLLGKRELGKE-----	314			
	NOV18b	301	LGRLLLGKRELGKE-----	314			
	gi 17445344	301	LGRLLLGKRELGKE-----	314			
	gi 5901478	237	-----	237			
	gi 13624329	301	LKKLMRFHHKSTKIKRNCKS	320			
30	gi 12054431	301	LKKLMRFHHKSTKIKRNCKS	320			
	gi 12054429	301	LKKLMRFHHKSTKIKRNCKS	320			

Tables 18G lists the domain descriptions from DOMAIN analysis results against NOV18. This indicates that the NOV18 sequence has properties similar to those of other proteins known to contain this domain.

Table 18G Domain Analysis of NOV18

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 95.1 bits (235), Expect = 5e-21

40	NOV18:	41	GNTTIIILVSRLDPHLHTPMYFFLAHLSFLDLSFTTSSIPQLLYNLNGCDKTISYMGCAIQ	100
	Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
	NOV18:	101	LFLFLGLGGVECLLLAVMAYDRCAICKPLHYMVMNPRLCRGLVSVTWGCGVANSLAMS	160
45	Sbjct:	61	GALFVVNGYASILLTIAISIDRYLAIVHPLRYRRIRTPRAKVLILLVWVLALLLSLP--	118
	NOV18:	161	PVTLRLPRCGHHEVDHFLREMPALIRMACVSTVAIEGTVFVLKKGVLSPLVFILLSSYSY	220
50	Sbjct:	119	PLLFVSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVL-----PLLVLVLCYTR	171
	NOV18:	221	IVRAV-----LQIRASGRQKAFGTGCGSHLTVVSLFYG----NIIYMYMQPGASSS	267
	Sbjct:	172	ILRTLKRARSQRSLSKRRSSSERKAAKMLLVVVVVVFLCWLPHYHIVLLDLSLCLLSIWRV	231
55	NOV18:	268	QDQGMFLMLFYNIIVTPLLNLPIY	290
			+ + + +	

Sbjct: 232 LPTALLITLWLAYVNSCLNPIIY 254

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV18 nucleic acid of the invention encoding a G-Protein Coupled Receptor -like protein includes the nucleic acid whose sequence is provided in Table 18A, 20C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 18A or 20C while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV18 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 18B or 20D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 18B or 20D while still encoding a protein that
5 maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 40 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

10 The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV18) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV18 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV18 nucleic acids and proteins of the invention are useful in potential
20 therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to
25 starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and
30 Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy(DRPLA) Hypophosphatemic rickets,

autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV18 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV18 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV18 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV19

NOV19 includes three novel G-Protein Coupled Receptor -like proteins disclosed below. The disclosed sequences have been named NOV19a and NOV19b.

NOV19a

A disclosed NOV19a nucleic acid of 1046 nucleotides (also referred to as CG56665-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 19A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 14-16 and ending with a TGA codon at nucleotides 1019-1021. The start and stop codons are shown in bold in Table 19A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 19A. NOV19a nucleotide sequence (SEQ ID NO:81).

TCAACATTATTACATGAACATTTTCAGATGTCATCTCCTTTTGATATTTTGGTTTCAGCCATGAAACAGGAAA
TCAAAGTTTTGGGACAGATTTTCTACTTGTGGTCTTTTCCAATATGGCTGGATAAACTCTCTTCTTTGT
CGTCATTGCCACCCTCTTTACAGTTGCTCTGACAGGAAATATCATGCTGATCCACCTCATTTCGACTGAACAC
CAGACTCCACACTCCAATGTACTTTCTGCTCAGTCAGCTCTCCATCGTTGACCTCATGTACATCTCCACCAC
AGTGCCCAAGATGGCAGTCAGCTTCCTCTCACAGAGTAAGACCATTAGATTTTTTGGGCTGTGAGATTCAAAC
GTATGTGTTCTTGGCCCTTGGTGGAAGTGAAGCCCTTCTCCTTGGTTTTATGTCTTATGATCGCTATGTAGC
TATCTGTCAACCCTTACATTATCCTATGCTTATGAGCAAGAAGATCTGCTGCCTCATGGTTGCATGTGCAATG
GGCCAGTGGTTCTATCAATGCTTTCATACATACATTGTATGTGTTTCAGCTTCCATTCTGTAGGTCTCGGCT
CATTAACCACTTTTCTGTGAAGTTCCAGCTCTACTATCATTTGGTGTGTCAGGACACCTCCCAGTATGAGTA
TACAGTCCTCCTGAGTGGACTTATTATCTTGCTACTACCATTCCTAGCCATTCTGGCTTCCTATGCTCGTGT
GCTTATTGTGGTATTCAGATGAGCTCAGGAAAAGGACAGGCAAAAGCTGTTTCCACTTGTTCTCCACCT
GATTGTGGCAAGCCTGTTCTATGCAACCACTCTCTTACCTACACAAGGCCACACTCCTTGCGTTCCTTC
ACGGGATAAGGCGTGGCAGTATTTTACACCATTTGCACACCTCTACTGAACCCATTTATCTACAGCCTGAG
AAATAAGGAAGTGACGGGGCAGTGAGGAGACTGTTGGATATTGGATATGCTGTAGAAAATATGACTTCAG
ATCTCTGTATTGATTGAGCATTAACAACATAAAAAGCT

The disclosed NOV19a polypeptide (SEQ ID NO:82) encoded by SEQ ID NO:81 has 335 amino acid residues and is presented in Table 19B using the one-letter amino acid code.

Table 19B. Encoded NOV19a protein sequence (SEQ ID NO:82).

MNISDVISFDILVSAMKTGNQSFQDFFLLVGLFYQGWINSLLFVVIATLFTVALTGNIIMLIHLI
RLNTRLHTPMYFLLSQLSIVDLMIYSTTVPKMAVSFSLSQSKTIRFLGCEIQTYVFLALGGTEAL
LLGFMSYDRYVAICHPLHYPMLMSKKICCLMVACAWASGSINAFIHTLYVFQLPFCRSRLINHF
FCEVPALLSLVCQDTSQYEYTVLLSGLI ILLLPFLAILASYARVLIVVFQMSGKGQAKAVSTC
SSHLIVASLFYATTLFTYTRPHSLRSPSRDKAVAVFYTIIVTPLLNPFIYSLRNKEVTGAVRRLLGYWIC
CRKYDFRSLY

A search of sequence databases reveals that the NOV19a amino acid sequence has 155 of 309 amino acid residues (50%) identical to, and 199 of 309 amino acid residues (64%) similar to, the 316 amino acid residue ptnr: TREMBLNEW-ACC:AAG45196 protein from *Mus musculus* (Mouse) (T2 Olfactory Receptor) ($E = 9.3e^{-79}$).

NOV19b

A disclosed NOV19b nucleic acid of 1046 nucleotides (also referred to as CG56665-02) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 19C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 59-60 and ending with a TGA codon at nucleotides 1019-1021. The start and stop codons are shown in bold in Table 19C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 19C. NOV19b nucleotide sequence (SEQ ID NO:83).

TCAACATTATTACATGAACATTTTCAGATGTCATCTCCTTTGATATTTTGGTTTCAGCCATGAAAAACAGGAAA
TCAAAGTTTTGGGACAGATTTTCTACTTGTGGTCTTTTTCCAATATGGCTGGATAAACTCTCTCTTTGT
CGTCATTGCCACCTCTTTACAGTTGCTCTGACAGGAAATATCATGCTGATCCACCTCATTCGACTGAACAC
CAGACTCCACACTCCAATGTACTTTCTGCTCAGTCAGCTCTCCATCGTTGACCTCATGTACATCTCCACCAC
AGTGCCCAAGATGGCAGTCAGCTTCTCTCACAGTAAGACCATTAGATTTTGGGCTGTGAGATTCAAAC
GATGTGTTCTTGGCCCTTGGTGGAAGTCCAGCTCTACTATCATTGATGTGTCAGGACACCTCCAGTATGAGTA
TATCTGTACACCTTTACATTATCCTATGCTTATGAGCAAGAAGATCTGCTGCCTCATGGTTGCATGTGCATG
GGCAGTGGTTCTATCAATGCTTTTACATACATTGTATGTGTTTCAGCTTCCATTCTGTAGGTCTCGGCT
CATTAACCACTTTTCTGTGAAGTTCAGCTCTACTATCATTGATGTGTCAGGACACCTCCAGTATGAGTA
TACAGTCCTCTGAGTGGACTTATTATCTTGCTACTACCATTCCTAGCCATTCTGGCTTCCATGCTCGTGT
GCTTATTGTGGTATTCCAGATGAGCTCAGGAAAAGGACAGGCAAAAGCTGTTTCCACTTGTTCCTCCACCT
GATTGTGGCAAGCCTGTTCTATGCAACCACTCTCTTACCTACACAAGGCCACACTCCTTGCCTTCCCTTC
ACGGGATAAGGCGGTGGCAGTATTTACACCATTTGTACACCTCTACTGAACCCATTTATCTACAGCCTGAG
AAATAAGGAAGTGACGGGGCAGTGAGGAGACTGTTGGGATATTGGATATGCTGTAGAAAATATGACTTCAG
ATCTCTGTATTGATTGAGCATTAAACAACATAAAAGCT

In a search of public sequence databases, the NOV19b nucleic acid sequence has 592 of 910 bases (65%) identical to a gb:GENBANK-ID:GGCOR4GEN|acc:X94744.1 mRNA from *Gallus gallus* (*G. gallus* cor4 DNA for olfactory receptor 4) ($E = 7.8e^{-48}$).

The disclosed NOV19b polypeptide (SEQ ID NO:84) encoded by SEQ ID NO:83 has 320 amino acid residues and is presented in Table 19D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV19b has A signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.4600. Alternatively,

NOV19b may also localize to the microbody (peroxisome) with a certainty of 0.2188, the endoplasmic reticulum (membrane) with a certainty of 0.1000, or in the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV19b is between positions 40 and 41: ALT-GN.

5

Table 19D. Encoded NOV19b protein sequence (SEQ ID NO:84).
MKTGNQSFQTDFLLVGLFQYQWINSLLFVVIATLFTVALTGNIIMLIHLIRLNTRLHTPMYFLLSQLSIVDLM YISTTVPKMAVSFLSQSKTIRFLGCEIQTYVFLALGGTEALLLGFMSSYDRYVAICHPLHYPLMSKKICCLM VACAWASGSINAFIHTLYVFLPFCRSRLINHFFCEVPALLSLMCQDTSQYEYTVLLSGLIILLPFLAILA SYARVLIIVFQMSSGKGQAKAVSTCSSHLIVASLFYATTLFTYTRPHSLRSPSRDKAVAVFYTIVTPLLNPF IYSLRNKEVTGAVRRLLGYWICCRKYDFRSLY

A search of sequence databases reveals that the NOV19b amino acid sequence has 155 of 306 amino acid residues (50%) identical to, and 198 of 306 amino acid residues (64%) similar to, the 316 amino acid residue ptnr:TREMBLNEW-ACC:BAB30304 protein from *Mus musculus* (Mouse) (Adult Male Testis cDNA, Riken Full-Length Enriched Library, Clone:4932441h21, Full Insert Sequence) ($E = 1.3e^{-79}$).

NOV19b is predicted to be expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV19a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 19E.

Table 19E. BLAST results for NOV19a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17445348 ref XP_060559.1 (XM_060559)	similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [<i>Homo sapiens</i>]	533	300/301 (99%)	301/301 (99%)	e-143
gi 17437056 ref XP_060314.1 (XM_060314)	similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [<i>Homo sapiens</i>]	695	169/310 (54%)	224/310 (71%)	5e-84
gi 17445356 ref XP_060561.1 (XM_060561)	similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [<i>Homo sapiens</i>]	312	172/305 (56%)	223/305 (72%)	3e-80
gi 17456595 ref XP_065073.1 (XM_065073)	similar to olfactory receptor (H. sapiens) [<i>Homo sapiens</i>]	638	142/292 (48%)	188/292 (63%)	7e-78
gi 17475192 ref XP_062796.1 (XM_062796)	similar to olfactory receptor (H. sapiens) [<i>Homo sapiens</i>]	315	154/299 (51%)	209/299 (69%)	2e-77

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 19F. In the ClustalW alignment of the NOV19 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 19F. ClustalW Analysis of NOV19

- 1) Novel NOV19a (SEQ ID NO:82)
- 2) Novel NOV19b (SEQ ID NO:84)
- 3) gi|17445348|ref|XP_060559.1| (XM_060559) similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:395)
- 4) gi|17437056|ref|XP_060314.1| (XM_060314) similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:397)
- 5) gi|17445356|ref|XP_060561.1| (XM_060561) similar to OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:394)

6) gi|17456595|ref|XP_065073.1| (XM_065073) similar to olfactory receptor (H. sapiens) [Homo sapiens] (SEQ ID NO:398)
 7) gi|17475192|ref|XP_062796.1| (XM_062796) similar to olfactory receptor (H. sapiens) [Homo sapiens] (SEQ ID NO:404)

5

			10	20	30	40	50	60
NOV19a	1	MNISDVISFDILVSAMKTGNQSF	TD	FL	LV	GL	FOYGWINSLLFVVI	TATLFTV
NOV19b	1	-----MKTGNQSF	TD	FL	LV	GL	FOYGWINSLLFVVI	TATLFTV
gi 17445348	1	-----MKTGNQSF	TD	FL	LV	GL	FOYGWINSLLFVVI	TATLFTV
gi 17437056	1	-----MCSGNQTSQ	NOTASTD	FL	IT	GL	FAESKHAALLYTV	TFLLFLMALTGNAIL
gi 17445356	1	-----MENYNQTS-	TD	FL	IL	GL	FPPSKIGLFL	FILFVLIFLMALIGNESM
gi 17456595	1	-----MGDVNQSVASD	FL	LV	GL	FSHSGSRQLL	FSLVAVMFVIGLL	LGNTVL
gi 17475192	1	-----METWVNQSY	TD	GF	LL	GI	FSHSTADLV	LFVSVMVAVETVALCGNVLL

15

			70	80	90	100	110	120
NOV19a	61	IHLIRINTRLHTPMYFLLS	QSL	STVD	LMYIST	TVPKMAVS	FLSQSKTIR	FLGCB
NOV19b	46	IHLIRINTRLHTPMYFLLS	QSL	STVD	LMYIST	TVPKMAVS	FLSQSKTIR	FLGCB
gi 17445348	46	IHLIRINTRLHTPMYFLLS	QSL	STVD	LMYIST	TVPKMAVS	FLSQSKTIR	FLGCB
gi 17437056	51	ILLIHSEPRLHTPMYF	FLS	QALMD	LMYLCV	TVPKMLVG	QVTGDDT	ISPSGCGI
gi 17445356	45	ILLIFLDTHLHTPMYFLLS	QSL	IDL	NYIST	TVPKMASD	FLYGNK	STISFICG
gi 17456595	46	LFLLRVDSRLHTPMYFLLS	QSL	SL	FDIG	CPMVIT	PKMASD	FLRGE
gi 17475192	47	IFLLIYMDPHLHTPMYFLLS	QSL	SLMD	LMV	CTNV	PKMAAN	FLSGRK

25

			130	140	150	160	170	180
NOV19a	121	ALCGTEALLLG	FMSYDR	VAIC	PHYP	MLMSKKIC	CLMVACAW	AGS
NOV19b	106	ALCGTEALLLG	FMSYDR	VAIC	PHYP	MLMSKKIC	CLMVACAW	AGS
gi 17445348	106	ALCGTEALLLG	FMSYDR	VAIC	PHYP	MLMSKKIC	CLMVACAW	AGS
gi 17437056	111	TLACAEVFL	LAAMAYDR	VAIC	PHYP	MLNQRV	COLV	SACW
gi 17445356	105	TFACAEALL	LTSMAYDR	VAIC	PHYP	TRMSK	RMVLMIT	GSMMIG
gi 17456595	106	LMGVAEGV	LLVLM	SYDR	VAV	COPL	CPV	LMR
gi 17475192	107	CLVCGSEGL	LLGLMAYDR	VAIC	PHYP	MLNQRV	COLQ	ITGSS

35

			190	200	210	220	230	240
NOV19a	181	LPFCRSRLIN	HFFCEV	PALL	SLVC	ODTS	QOYE	TVLL
NOV19b	166	LPFCRSRLIN	HFFCEV	PALL	SLVC	ODTS	QOYE	TVLL
gi 17445348	166	LPFCRSRLIN	HFFCEV	PALL	SLVC	ODTS	QOYE	TVLL
gi 17437056	171	FPFCQSRK	ILSFFC	ET	PALL	KLK	SCD	VS
gi 17445356	165	IPYCKSR	AINHFF	CEV	PALL	TLAC	DTWV	VEY
gi 17456595	166	FPYCASR	IVHFF	CEV	PALL	KLSC	ADT	CAY
gi 17475192	167	FPYCGLR	KVNHFF	CEML	SL	KLAC	VT	SL

45

			250	260	270	280	290	300
NOV19a	241	FOMSSGKGQ	AKAV	STC	SSH	LIVAS	LFY	AT
NOV19b	226	FOMSSGKGQ	AKAV	STC	SSH	LIVAS	LFY	AT
gi 17445348	226	FOMSSGKGQ	AKAV	STC	SSH	LIVAS	LFY	AT
gi 17437056	231	HRMNSA	ACH	RKAL	ATC	SSH	MI	IV
gi 17445356	225	YRMHSA	EGR	KAY	STC	SH	LT	TV
gi 17456595	226	LSMRSE	EAR	KA	VT	C	SSH	ET
gi 17475192	227	LQMHS	AQAW	KAL	ATC	SSH	LT	AV

55

			310	320	330	340	350	360
NOV19a	301	NPTIYSLRN	KEV	TG	AVR	-----	-----	-----
NOV19b	286	NPTIYSLRN	KEV	TG	AVR	-----	-----	-----
gi 17445348	286	NPTIYSLRN	KEV	TG	AM	ASD	FL	SG
gi 17437056	291	NPLIYSLRN	KEV	TR	AMR	SMM	QAME	Q
gi 17445356	285	NPTIYSLRN	KEV	M	GA	IT	-----	-----
gi 17456595	286	NPLIYSLRN	PE	SN	AN	HR	OP	P
gi 17475192	287	NPLIYSLRN	RE	V	M	GA	LR	-----

65

			370	380	390	400	410	420
NOV19a	317	-----	-----	-----	-----	-----	-----	-----
NOV19b	302	-----	-----	-----	-----	-----	-----	-----

	gi 17445348	311	-----ISFTGCG	318
	gi 17437056	351	IASNVVKIILIHDSRLHTPMYFLLSQLSLRDILYISTIVPKMLVDQVMSQR AISFAGCT	410
	gi 17445356	301	-----	301
	gi 17456595	332	-----ESHVSLISLVE	343
5	gi 17475192	303	-----	303
			430 440 450 460 470 480	
			
	NOV19a	317	-----RILGYWTC-----	325
10	NOV19b	302	-----RILGYWTC-----	310
	gi 17445348	319	IQSFFFSALGGAEALLASMA YDRYIAICFPLHYP IRMSKRMCVLMITGSWTIGSINACA	378
	gi 17437056	411	AQHFLYLTLTAGAEFFLLGLMSYDRYVAICNPLHY PVLSMRKICWLIVAAAWLGGSIDGFL	478
	gi 17445356	301	-----RVIQNIIFS-----	309
	gi 17456595	344	PPAVEVVTGASVKGCPRTWCLPREQVLWDGPDSGT SLESKP HQEGLSDMHHSNTICTLV	403
15	gi 17475192	303	-----KGLDRCRTG-----	312
			490 500 510 520 530 540	
			
	NOV19a	325	-----CRKYDFRSLY-----	335
20	NOV19b	310	-----CRKYDFRSLY-----	320
	gi 17445348	379	HTV----YVLHIPYCQSRAINHF FCDVPAMVTLACMDT WVYE GTVF LSTTIF LVFPFIA	433
	gi 17437056	471	LTP----VTMQPFPCASREINHFFCE VPALLKL SCTDT SAYETAMY VCCIMMLLI PFSV	525
	gi 17445356	309	-----VKM-----	312
	gi 17456595	404	SELNQFWAYPIQHDLPEVLLTPAPCKVGAI IHPAAREDTLNTSQETPGTPKCYRGKNI	463
25	gi 17475192	312	-----SQH-----	315
			550 560 570 580 590 600	
			
	NOV19a	335	-----	335
30	NOV19b	320	-----	320
	gi 17445348	434	ISCSYGRVLLAVYHMKS AEGRKKAYLT CSTHLTV VTFFYAPFVYTYLRPRS LRSPTE DKV	493
	gi 17437056	526	ISGSYTRILITVYRMSEA EGRG KAVATC SSHMVVVS LFYGAAMY TYVLP HS YHTPEQDKA	585
	gi 17445356	312	-----	312
	gi 17456595	464	KGVKEGKAEP EGPV GPET VGSKTEMNFAGSE FKEVNFR CTASMHN-- SPDVTSDPVLQAA	521
35	gi 17475192	315	-----	315
			610 620 630 640 650 660	
			
	NOV19a	335	-----	335
40	NOV19b	320	-----	320
	gi 17445348	494	LAVFYTTLTPLMNPIIYS LRNKEVMGALT RV SQRICSGKM----	533
	gi 17437056	586	VSAFYTILTPLMNPLIIYS LRNKDV T GALQKV VGRMEWK TL PFQALQVR CVKWRRSVLVSS	645
	gi 17445356	312	-----	312
	gi 17456595	522	MDVGFGSGLPDVS QSHSKTLWGARGRGPSIR RQREFMPEEK--- KD TVYWEKR RKNNNEAAK	578
45	gi 17475192	315	-----	315
			670 680 690 700 710 720	
			
	NOV19a	335	-----	335
50	NOV19b	320	-----	320
	gi 17445348	533	-----	533
	gi 17437056	646	FIATERTLADTSHSSSHA EFPERGVR MN---- C SKLF SLVE EPVTSL GD LFNFR----	695
	gi 17445356	312	-----	312
	gi 17456595	579	RSREKRRRLNDAAIE GR LAALMEEN ALLKGELKALKLR FG LLPLTG SAIG SPLDW GPPAWG	638
55	gi 17475192	315	-----	315
	NOV19a	335	335	
60	NOV19b	320	320	
	gi 17445348	533	533	
	gi 17437056	695	695	
	gi 17445356	312	312	
	gi 17456595	638	638	
65	gi 17475192	315	315	

Table 19G lists the domain description from DOMAIN analysis results against NOV19. This indicates that the NOV19 sequence has properties similar to those of other proteins known to contain this domain.

Table 19G Domain Analysis of NOV19

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 91.3 bits (225), Expect = 8e-20

5

NOV19:	56	GNIMLIHLIRLNLRLHTPMYFLLSQLSIVDLMYISTTVPKMAVSFLSQSKTIRFLGCEIQ	115
		+++ + + ++ +++ + ++	
Sbict:	1	GNLLVILVILRTKKLRPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60

10

NOV19: 116 TYVFLALGGTEALLLGFMFSYDRYVAICHPLHYPMLSKKICCLMVACAWASGSINAFIHT 175
+|+ | || +|||+|| ||| | + + + +++ | + +
Sbjct: 61 GALTFFVNGYASILLTALSIDRYLAIVHPLRYRRIPTPRRAKVLILLVWVLALLSLPPL 120

15

NOV19:	176	LYVQQLPFCRSRLINHFFCEVPALLSLVCQDTSQYEYTVLLSGLIILLPLFLAILASYAR	235
		+ + + + + ++ +	
Sbjct:	121	LFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPPLVLVLCYTRILRTLKRRA	180

20

NOV19: 236 VLIVVFQMSSGKGQAKAVSTCSSHLIVASLFY---ATTLFYTRPHSLRSPSRDKAVAV 291
 + | + | ++ + | + | + +
 Sbjct: 181 RSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSCLLSIWRVLPALTITL 240

25

```
NOV19:  292  FYTIVTPLLNPFIY  305
          +   |   |||  ||
Sbjct:  241  WLAYVNSCLNPFIY  254
```

G-Protein Coupled Receptor (GPCRs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of various signals. Previously, GPCR genes cloned in different species were from random locations in the respective genomes. The human GPCR genes are intron less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

35

Olfactory receptors (ORs) have been identified as extremely large subfamily of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Previously, OR genes cloned in different species were from random locations in the respective genomes. The human OR genes are intron less and belong to four different gene subfamilies,

displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV19 nucleic acid of the invention encoding a G-Protein Coupled Receptor -like protein includes the nucleic acid whose sequence is provided in Table 19A, 19C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 19A or 19C while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 35 percent of the bases may be so changed.

The disclosed NOV19 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 19B or 19D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 19B or 19D while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 52 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV19) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV19 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene

delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV19 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluyasian atrophy(DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV19 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV19 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV19 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV20

A disclosed NOV20 nucleic acid of 1027 nucleotides (also referred to as CG56665-01) encoding a novel G-Protein Coupled Receptor -like protein is shown in Table 20A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TAG codon at nucleotides 940-942. The start and stop codons are shown in bold in Table 20A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 20A. NOV20 nucleotide sequence (SEQ ID NO:85).

ATGATCTGCTCAGCTATCAACCTACACTTACTACTGGCAGTTAAGATGATTACCCCTGTCTGGATTCTTGCT
CCTCGGGAGCAAGGGCTGTTCTGCTGATTATCTGCGAGTGTGGTGGGGAACCTGCTCATCATTGCAGTC
ATCACTCTCGATCAGCATCTTACACACCCATGTAATCTTCTGCTGAAGAACCTCTCCGTTTTGGATCTGTGC
TACATCTCAGTCACTGTGCCTAAATCCATCCGTAACTCCCTGACTCGCAGAAGCTCCATCTTATCTTGGC
TGTGTGGCTCAAGTCTATTTTTCTCTGCTTTGTCATCTGCTGAGCTGGCCTTCCTTACTGTCATGTCTTAT
GACCGCTATGTTGCCATTGGCCACCCCTCCAATACAGAGCCGTGATGACATCAGGAGGGTGTATCAGATG
GCAGTCACCACCTGGCTAAGCTGCTTTTCTACGCAGCCGTCCACACTGGCAACATGTTTCGGGAGCACGTT
TGCAGATCCAGTGTGATCCACCACTTCTCCGTGACATCCCTCATGTGTTGGCCCTGGTTTCTGTGAGGTT
TTCTTTGTAGAGTTTTTGACCCTGGCCCTGAGCTCATGCTTGGTTCTGGGATGCTTTATTCTCATGATGATC
TCCTATTTCCAAATCTTCTCAACGGTGTCTCAGAATCCCTTCAGGACAGAGTCGAGCAAAAGCCTTCTCCACC
TGCTCCCCCAGCTCATTGTCTCATGCTCTTTCTTACCACAGGGCTCTTTGCTGCCTTAGGACCAATTGCA
AAAGCTCTGTCCATTTCAGGATTTAGTGATTGCTCTGACATACACAGTTTGCCTCCCTTCCATCCCATC
ATATATAGTCTTAGGAATAAGGAGATTAAACAGCCATGTGGAGACTCTTGTGAAGATATATTTCTGCAA
AAGTAGAACATCCTGGTCTTACTATAGAAGATCTGCAACAAAACCCCAAAAAGCATAAATACTTTATGAC
AAAAAAGATGAAAAATT

The disclosed NOV20 polypeptide (SEQ ID NO:86) encoded by SEQ ID NO:85 has 313 amino acid residues and is presented in Table 20B using the one-letter amino acid code.

Table 20B. Encoded NOV20 protein sequence (SEQ ID NO:86).

MICSAINLHLLAVKMIHPVWILAPREQGLFLLIYLAFLVGNLLIIIAVITLDQHLHTPMYFFLK
NLSVLDLCYISVTVPKSIRNSLTRSSISYLGCVAVQVFFSAFASAEALFLTVMSYDRYVAICH
PLQYRAVMTSGGCYQMAVTTWLSFCFSYAAVHTGNMFREHVCRSSVIHQFFRDIPHVLALVSCEV
FFVEFLTLALSSCLVLGCFILMMISYFQIFSTVLRIPSGQSRKAFSTCSPQLIVIMLFLTGL
FAALGPIAKALSIQDLVIALTYTVLPPFLNPIIYSLRNKEIKTAMWRLFVKIYFLQK

A search of sequence databases reveals that the NOV20 amino acid sequence has 134 of 278 amino acid residues (48%) identical to, and 179 of 278 amino acid residues (64%) similar to, the 321 amino acid residue ptrn: SPTREMBL-ACC:Q9UGF5 BA150A6.4 protein from *Homo sapiens* (Human) (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) ($E = 2.4e^{-64}$).

The disclosed NOV20 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 20C.

Table 20C. BLAST results for NOV20

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17437075 ref XP_060319.1 (XM_060319)	similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>]	311	287/294 (97%)	288/294 (97%)	e-134

gi 17445373 ref XP_060567.1 (XM_060567)	similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>]	309	147/272 (54%)	188/272 (69%)	8e-63
gi 17445394 ref XP_060572.1 (XM_060572)	similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>]	316	133/283 (46%)	187/283 (65%)	2e-61
gi 17437015 ref XP_060307.1 (XM_060307)	similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>]	312	139/291 (47%)	189/291 (64%)	9e-59
gi 17464351 ref XP_069462.1 (XM_069462)	similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>]	321	133/278 (47%)	175/278 (62%)	3e-57

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 20D. In the ClustalW alignment of the NOV20 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 20D. ClustalW Analysis of NOV20

1) Novel NOV20 (SEQ ID NO:86)	
3) gi 17437075 ref XP_060319.1 (XM_060319) similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:405)	
4) gi 17445373 ref XP_060567.1 (XM_060567) similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:406)	
5) gi 17445394 ref XP_060572.1 (XM_060572) similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:407)	
6) gi 17437015 ref XP_060307.1 (XM_060307) similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:408)	
7) gi 17464351 ref XP_069462.1 (XM_069462) similar to OLFACTORY RECEPTOR 5U1 (HS6M1-28) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:409)	
	10 20 30 40 50 60
NOV20 1	-----MTCSAINTHLLLAVKMIHPVWILAPREQGLFLLNYLAVLVGNLLIITAVIT 50
gi 17437075 1	-----MDNLTKVTEFLLMEFSG-IWELOVLHAGLFLLIYLAFLVGNLLIITAVIT 48
gi 17445373 1	-----MANLTIIVTEFLLMGFST-NKNMCIILHSILFLLIYLCALMGNVLIITMTT 48
gi 17445394 1	MKAGRNGTTTSDRNOTEVAEFELMGFSN-SWDIOTVHAALFELVYLAAVIGNLLIITTT 59
gi 17437015 1	-----MPNSTTVMEFLLMRFSQ-VWTLQILHSASFEMLYLVLTMGNLIIVTMTT 48
gi 17464351 1	-----MVNLTSMGFLLMGFSD-ERKLQILHALVFLVLYLLALVGNLLIITTTT 48
	70 80 90 100 110 120



40

45

Table 20E Domain Analysis of NOV20

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family)
(SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 83.6 bits (205), Expect = 2e-17

55

5
10
10

```

NOV20: 161 GNMFRHVCRRSSVIHQFFRDIPHVLALVSCVFFVEFLTLALSSCLVLGCFILMMISYFQ 220
          +   |   +           +   +   +   |   |   |   |   +
Sbjct: 121 LFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLP LLVILVCYTRILRTLKRRA 180

NOV20: 221 IFSTVLRIPSGQSRKAFSTCSPQLIVIMLFLTTLGLFAALGPIAKALSIQDLVIALT--- 277
          | + | | |           ++ ++ + |   +   |   +   +   |   |
Sbjct: 181 RSQRSLKRRSSSERKAAKMLLVVVVFVLCWLPYHIVLLLDLCLLSIWRVLP TALLITL 240

NOV20: 278 -YTVLPPFLNP I I Y 290
          +   | | | | |
Sbjct: 241 WLAYVNSCLNP I I Y 254

```

15 G-Protein Coupled Receptor (GPCRs) have been identified as extremely large
subfamily of G protein-coupled receptors in a number of species. These receptors share a
seven transmembrane domain structure with many neurotransmitter and hormone receptors,
and are likely to underlie the recognition and G-protein-mediated transduction of various
signals. Previously, GPCR genes cloned in different species were from random locations in the
respective genomes. The human GPCR genes are intron less and belong to four different gene
subfamilies, displaying great sequence variability. These genes are dominantly expressed in
olfactory epithelium.

25 Olfactory receptors (ORs) have been identified as extremely large subfamily of G
protein-coupled receptors in a number of species. These receptors share a seven
transmembrane domain structure with many neurotransmitter and hormone receptors, and are
likely to underlie the recognition and G-protein-mediated transduction of odorant signals.
Previously, OR genes cloned in different species were from random locations in the respective
genomes. The human OR genes are intron less and belong to four different gene subfamilies,
displaying great sequence variability. These genes are dominantly expressed in olfactory
epithelium.

30 The disclosed NOV20 nucleic acid of the invention encoding a G-Protein Coupled
Receptor -like protein includes the nucleic acid whose sequence is provided in Table 20A or a
fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose
bases may be changed from the corresponding base shown in Table 20A while still encoding a
protein that maintains its G-Protein Coupled Receptor -like activities and physiological
35 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids
whose sequences are complementary to those just described, including nucleic acid fragments
that are complementary to any of the nucleic acids just described. The invention additionally
includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures
include chemical modifications. Such modifications include, by way of nonlimiting example,
40 modified bases, and nucleic acids whose sugar phosphate backbones are modified or
derivatized. These modifications are carried out at least in part to enhance the chemical

stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

The disclosed NOV20 protein of the invention includes the G-Protein Coupled Receptor -like protein whose sequence is provided in Table 20B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 20B while still encoding a protein that maintains its G-Protein Coupled Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 54 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G-Protein Coupled Receptor -like protein (NOV20) is a member of a "G-Protein Coupled Receptor family". Therefore, the NOV20 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV20 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy(DRPLA) Hypophosphatemic rickets,

autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV20 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV20 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV20 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV21

NOV21 includes three novel adrenal secretory serine protease-like proteins disclosed below. The disclosed sequences have been named NOV21a and NOV21b.

NOV21a

A disclosed NOV21a nucleic acid of 1028 nucleotides (also referred to as CG56639-01) encoding a novel adrenal secretory serine protease-like protein is shown in Table 21A. An open reading frame was identified beginning with an TCG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 769-771. The start and stop codons are shown in bold in Table 21A, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon of NOV21a is not a traditional initiation codon, NOV21a could be a partial reading frame that extends further in the 5' direction.

Table 21A. NOV21a nucleotide sequence (SEQ ID NO:87).

TCGCCATTTCAGACGCCCCGGAGGCCACCACACACCCAGCTACCAGACTGTGGCCTGGCGCCGGCCGCG
CTCACCAGGATTGTGGGCGGCAGCGCAGCGGGCCGTGGGGAGTGGCCGTGGCAGGTGAGCCTGTGGCTGCGG
CGCCGGGAACACCGTTGCGGGGCCGTGCTGGTGGCAGAGAGGTGGCTGCTGTGCGCGCGCACTGCTTCGAC
GTCTACGGGGACCCCAAGCAGTGGGCGGCCCTTCCTAGGCACGCCGTTCTGAGCGGCGCGGAGGGGCGAGCTG
GAGCGCGTGGCGCGCATCTACAAGCACCCGTTCTACAATCTCTACACGCTCGACTACGACGTGGCGCTGCTG
GAGCTGGCGGGGCCGTTGCGTTCGAGCCGCTGGTGGCTCCCATCTGCGCTGCCCCAGCCCGCGCGCAGCC
CCGACGGGCACGCGCTGCGTCATCACCGGCTGGGGCTCGGTGCGCGAAGGAGGCTCCATGGCGCGGCGAGCTG
CAGAAGGCGGCCGTGCGCCTCCTCAGCGAGCAGACCTGCCGCCGCTTCTACCCAGTGCAGATCAGCAGCCGC
ATCTCTGAACCCCTTTCTTCTCTCCCCAACAGGGTGACGCTGGGGGACCCCTGGCCTGCAGGGAGCCCTCT
GGACGGTGGGTGCTAACTGGGTCACTAGCTGGGGCTATGGCTGTGGCCGGCCCCACTTCCCAGGTGTCTAT
ACCCGGGTGGCAGCTGTGAGAGGCTGGATAGGACAGACATCCAGGAGTGACCAACACGTGACTGCCAGGC
CGAGACTCTACGTGAAAGCAACAGGAGCAGAGGCCACCCAAACACCCACCCACCGTACCCTACCCAAGGA
CGGGTGTGGGGGGCTGTGGGTGATGGGGATGCATTTTGGTACCACCCCTTTGTTCCAATAAACACAGCCCCC
CCACCCTAGCTCACTGGCTCAGCACCTCAGTGTACAGCGAGGACCACCTGCCTGGTGCTTCACCAGGACCC
GGGGTGGAAACGAACAACCC

In a search of public sequence databases, the NOV21a nucleic acid sequence, located on chromosome 19, has 296 of 466 bases (63%) identical to a gb:GENBANK-ID:E13204|acc:E13204.1 mRNA from *Homo sapiens* (Human cDNA encoding a serine protease) ($E = 3.9e^{-18}$).

5 The disclosed NOV21a polypeptide (SEQ ID NO:88) encoded by SEQ ID NO:87 has 256 amino acid residues and is presented in Table 21B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV21a has A signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.7480. Alternatively, NOV21a may also localize to the lysosome (lumen) with a certainty of 0.3168, or the
10 mitochondrial matrix space with a certainty of 0.1000. The most likely cleavage site for NOV21a is between positions 68 and 69: SAA-HC.

Table 21B. Encoded NOV21a protein sequence (SEQ ID NO:88).

SPFPDAPEATHTQLPDCGLAPAALTRIVGSSAAGRGWEPQVSLWLRREHRCGAVLVAERWLLSAAHCFD VYGDPKQWAAFLGTPFLSGAEGQLERVARIYKHFFYNLYTLDYDVALLELAGPVRRSRLVRPICLPEPAPRP PDGTRCVITGWGSVREGGSMARQLQKAAVRLLEQTCCRFPVQISSRISEPPFFSPQQGDAGGPLACREPS GRVLTGVTSWGYGCGRPFGVYTRVAAVRGWIGQHIQE
--

A search of sequence databases reveals that the NOV21a amino acid sequence has 99
15 of 250 amino acid residues (39%) identical to, and 134 of 250 amino acid residues (53%) similar to, the 279 amino acid residue ptnr:SPTREMBL-ACC:Q9QZ74 protein from *Rattus norvegicus* (Rat) (Adrenal Secretory Serine Protease Precursor) ($E = 1.5e^{-42}$).

NOV21a is predicted to be expressed in at least the following tissues: Ovary, kidney, breast, lung, muscle, liver, spleen, blood, lymphocyte. This information was derived by
20 determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV21b

In the present invention, the target sequence identified previously, NOV21a, was
25 subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of
30 the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein

sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated Accession Number NOV21b. This differs from the previously identified sequence (NOV21a) in being a splice variant and a mature protein starting with serine.

A disclosed NOV21b nucleic acid of 785 nucleotides (also referred to as CG56639-02) encoding a novel adrenal secretory serine protease-like protein is shown in Table 21C. An open reading frame was identified beginning with an CTT initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 783-785. The start and stop codons are shown in bold in Table 21C, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon of NOV21b is not a traditional initiation codon, NOV21b could be a partial reading frame that extends further in the 5' direction.

Table 21C. NOV21b nucleotide sequence (SEQ ID NO:89).

CTTCGCCATTTCAGACGCCCCGGAGGCCACACACACCCAGCTACCAGACTGTGGCCTGGCGCCGCGCG
CGCTCACCAGGATTGTGGGCGGCAGCGCAGCGGGCCGTGGGGAGTGGCCGTGGCAGGTGAGCCTGTGGCTGC
GGCGCCGGGAACACCGTTGCGGGGCGGTGCTGGTGCCAGAGAGGTGGCTGCTGTCGGCGGCGCACTGCTTCG
ACGTCTACGGGGACCCCAAGCAGTGGGCGGCCCTTCCTAGGCACGCCGTTCTTGAGCGGCGCGGAGGGGCGAGC
TGGAGCGCGTGGCGCGCATCTACAAGCACCCGTTCTACAATCTCTACACGCTCGACTACGACGTGGCGCTGC
TGGAGCTGGCGGGGCGGTGCGTGCAGCCGCTGGTGGCTCCCATCTGCCTGCCCGAGCCCGCGCCGCGAC
CCCCGGACGGCACGCGCTGCGTCACTACCGGCTGGGGTCCGGTGC CGGAAGGAGGCTCCATGGCGCGGCAGC
TGCAGAAGGCGGCGGTGCGCCTCCTCAGCGAGCAGACCTGCCACCGCTTCTACCCAGTGCAGATCAGCAGCC
GCATGCTGTGTGCCGGCTTCCCGCAGGGTGGCGTGGACAGCTGCTCGGGTGACGCTGGGGGACCCCTGGCCT
GCAGGGAGCCCTCTGGACGGTGGGTGCTAACTGGGGTCACTAGCTGGGGCTATGGCTGTGGCGGCCCCACT
TCCCAGGTGTCTATACCCGGTGGCAGCTGTGAGAGGCTGGATAGGACAGCACATCCAGGAGTGA

In a search of public sequence databases, the NOV21b nucleic acid sequence, located on chromosome 19, has 160 of 162 bases (98%) identical to a gb:GENBANK-

ID:HUMLAMBBB|acc:M94363.1 mRNA from *Homo sapiens* (Human lamin B2 (LAMB2) gene and ppv1 gene sequence) ($E = 4.3e^{-59}$).

The disclosed NOV21b polypeptide (SEQ ID NO:90) encoded by SEQ ID NO:89 has 260 amino acid residues and is presented in Table 21D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV21b has A signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.7480. Alternatively, NOV21b may also localize to the lysosome (lumen) with a certainty of 0.3082, or the mitochondrial matrix space with a certainty of 0.1000. The most likely cleavage site for NOV21b is between positions 68 and 69: SAA-HC.

Table 21D. Encoded NOV21b protein sequence (SEQ ID NO:90).

SPFPDAPEATTHTQLPDCGLAPAALTRIVGSSAAGRGEPWQVSLWLRREHRCGAVLVAERWLLSAAHCFD
VYGDPKQWAAFLGTPFLSGAEGQLERVARIYKHFFYNLYTLDYDVALLELAGPVRRSRLVRPICLPEPAPRP
PDGTRCVITGWGSGVREGGSMARQLQKAAVRLLEQTCHRFYFPVQISSRMLCAGFPQGGVDSCSGDAGGPLAC
REPSGRWVLTGVTSWGYGCGRPHFPGVYTRVAAVRGWIGQHIQE

A search of sequence databases reveals that the NOV21b amino acid sequence has 123 of 250 amino acid residues (49%) identical to, and 154 of 250 amino acid residues (61%) similar to, the 855 amino acid residue ptnr:SPTREMBL-ACC:Q9Y5Y6 protein from *Homo sapiens* (Human) (Matriptase) ($E = 3.5e^{-59}$).

NOV21b is predicted to be expressed in at least the following tissues: adrenal gland, Ovary, kidney, breast, lung, muscle, liver, spleen, blood, lymphocyte. .

The disclosed NOV21a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 21E..

Table 21E. BLAST results for NOV21a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12836503 dbj BAB 23684.1 (AK004939)	data source:SPTR, source key:095519, evidence:ISS-homo log to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT)-putati ve [<i>Mus musculus</i>]	799	118/244 (48%)	153/244 (62%)	7e-55
gi 10257390 gb AAG1 5395.1 AF057145_1 (AF057145)	serine protease TADG15 [<i>Homo sapiens</i>]	855	115/250 (46%)	146/250 (58%)	6e-52

gi 11415040 ref NP_068813.1 (NM_021978)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [<i>Homo sapiens</i>]	855	115/250 (46%)	146/250 (58%)	7e-52
gi 7363445 ref NP_35306.2 (NM_011176)	protease, serine, 14 (epithin) [<i>Mus musculus</i>]	855	115/250 (46%)	144/250 (57%)	8e-52
gi 16758444 ref NP_446087.1 (NM_053635)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [<i>Rattus norvegicus</i>]	855	112/247 (45%)	141/247 (56%)	1e-51

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 21F. In the ClustalW alignment of the NOV21 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 21F. ClustalW Analysis of NOV21

- 1) Novel NOV21a (SEQ ID NO:88)
- 2) Novel NOV21b (SEQ ID NO:90)
- 3) gi|12836503|dbj|BAB23684.1| (AK004939) data source:SPTR, source key:O95519, evidence:ISS-homolog to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT)-putative [*Mus musculus*] (SEQ ID NO:410)
- 4) gi|10257390|gb|AAG15395.1|AF057145_1 (AF057145) serine protease TADG15 [*Homo sapiens*] (SEQ ID NO:411)
- 5) gi|11415040|ref|NP_068813.1| (NM_021978) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [*Homo sapiens*] (SEQ ID NO:412)
- 6) gi|7363445|ref|NP_035306.2| (NM_011176) protease, serine, 14 (epithin) [*Mus musculus*] (SEQ ID NO:413)
- 7) gi|16758444|ref|NP_446087.1| (NM_053635) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [*Rattus norvegicus*] (SEQ ID NO:414)

		10	20	30	40	50	60							
													
NOV21a	1	-----												1
NOV21b	1	-----												1
gi 12836503	1	MPTEVPQAAADGGQDAGDGE-----AAPEGKFKPKNTKR--KNRD--NVRFTP	47											
gi 10257390	1	MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVLA	60											
gi 11415040	1	MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVLA	60											
gi 7363445	1	MGSNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNVKKVEKHGPRRWVLA	60											
gi 16758444	1	MGNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNVKKVEKHGPRRWVLA	60											
		70	80	90	100	110	120							
													

	NOV21a	1	-----	1			
	NOV21b	1	-----	1			
5	gi 12836503	453	-----VN-GLCVP---ACDGTKDCPNGLDERNCVCR	480			
	gi 10257390	467	KELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCCKPLFWVCDSVNDCCGDNSEOGCSCP	526			
	gi 11415040	467	KELRCDGWADCTDHSDELNCSCDAGHQFTCKNKFCCKPLFWVCDSVNDCCGDNSEOGCSCP	526			
	gi 7363445	467	KELRCDGWADCPDYSDEYRCRCNATHQFTCKNQFCCKPLFWVCDSVNDCCGDSDEEGCSCP	526			
	gi 16758444	467	KDLRCDGWADCPDYSDEYRCRCNATHQFMCKNQFCCKPLFWVCDSVNDCCGDSDEEGCSCP	526			
10		550	560	570	580	590	600
	NOV21a	1	-----	SPFPDAPEATTHT	-----	QLP	16
	NOV21b	1	-----	SPFPDAPEATTHT	-----	QLP	16
15	gi 12836503	481	AMFQCQEDSTCTSLPRVCDROPDCNGSDEEQOEG--VPCGTFTFCEDRSCVKKPNPE	538			
	gi 10257390	527	AQTFRCNSNGKCLSKSQCCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPE	586			
	gi 11415040	527	AQTFRCNSNGKCLSKSQCCNGKDDCGDGSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPE	586			
	gi 7363445	527	AGSFKCSNGKCLPQSQCCKNGKDDCGDGSDEASCDNVNAVSTCKYTYRCQNGLCLSKGNPE	586			
	gi 16758444	527	AGSFKCSNGKCLPQSQCCKNGKDDCGDGSDEASCDNVNAVSTCKYTYRCQNGLCLSKGNPE	586			
20		610	620	630	640	650	660
	NOV21a	16	-----	DCGLAP-AALTRIVGSSAAGRGEWPQVSLWLRRRHRRCGAVLV	59		
	NOV21b	16	-----	DCGLAP-AALTRIVGSSAAGRGEWPQVSLWLRRRHRRCGAVLV	59		
25	gi 12836503	539	CDGQSDCRDGSDEQHCDCLGQ--LSSRIVGGTVSSEGEWPQASLQIRGR-HICCGALI	595			
	gi 10257390	587	CDGKEDCSDGSDCKDCDCGLRSFTQARVVGGTDADEGEWPQVSLHALCQGHICGASLI	646			
	gi 11415040	587	CDGKEDCSDGSDCKDCDCGLRSFTQARVVGGTDADEGEWPQVSLHALCQGHICGASLI	646			
	gi 7363445	587	CDGKTDSCDGSDCKDCGLRSFTQARVVGGTNADEGEWPQVSLHALCQGHICGASLI	646			
	gi 16758444	587	CDGKTDSCDGSDCKDCGLRSFTQARVVGGTNADEGEWPQVSLHALCQGHICGASLI	646			
30		670	680	690	700	710	720
	NOV21a	60	AERWLLSAAHCFD-----VYGDPKOWAAFLGTPFLS--GAEQCQLER-VARIYKHFFYNLY	111			
	NOV21b	60	AERWLLSAAHCFD-----VYGDPKOWAAFLGTPFLS--GAEQCQLER-VARIYKHFFYNLY	111			
35	gi 12836503	596	ADRWVITAAHCFQED---SMASPKLWTVFLGKMCRNS-RWPGEVSFVYSRLFLHPYHEED	651			
	gi 10257390	647	SPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQSQRSAFGVQERRLKRIISHPFFNDE	706			
	gi 11415040	647	SPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQSQRSAFGVQERRLKRIISHPFFNDE	706			
	gi 7363445	647	SPDWLVSAAHCFQDDKNFKYSDYTMWTAFLGLLDQSKRSASGVQELKLRKIITHPSFNDE	706			
	gi 16758444	647	SPDWLVSAAHCFQDETIFKYSDDHTMTAFLGLLDQSKRSASGVQEHKLRKIITHPSFNDE	706			
40		730	740	750	760	770	780
	NOV21a	112	TLDYDVALLELAGFVRRSRIVRPICLPEPAPRPEDGTRCVITGWGSVREGGSMARQLQKA	171			
	NOV21b	112	TLDYDVALLELAGFVRRSRIVRPICLPEPAPRPEDGTRCVITGWGSVREGGSMARQLQKA	171			
45	gi 12836503	652	SHDYDVALLOLDHFPVVYSATVRPVCLEARSHFEPGQHCWITGWGAQREGGSPVSNLQKV	711			
	gi 10257390	707	TFDYDIALLELEKPAEYSSMVRPICLEDAHVFPAGKAIWVTGWGHTQYGGTGALILQKG	766			
	gi 11415040	707	TFDYDIALLELEKPAEYSSMVRPICLEDAHVFPAGKAIWVTGWGHTQYGGTGALILQKG	766			
	gi 7363445	707	TFDYDIALLELEKSVESYSTVVRPICLEDAHVFPAGKAIWVTGWGHTKEGGTGALILQKG	766			
	gi 16758444	707	TFDYDIALLELEKPAEYSTVVRPICLEDNTHVFPAGKAIWVTGWGHTKEGGTGALILQKG	766			
50		790	800	810	820	830	840
	NOV21a	172	AVRLLESEQTCRRFYFVQIISRRISEPPHFS---PQCGDAGGPLACREPSGRWVLTGVTSW	227			
	NOV21b	172	AVRLLESEQTCRRFYFVQIISRRLCAGFPQGGVDSGSGDAGGPLACREPSGRWVLTGVTSW	231			
55	gi 12836503	712	DVQLVPDCLCEAYRYQVSPRMLCAGYRKGGKDACQGDSSGGLVCREPSGRWFLAGLVSW	771			
	gi 10257390	767	EIRVINQTTTCENLLPQQTIPRMMCVGFLSGGVDSQCGDSGGPLSSVEADGRIFOAGVVS	826			
	gi 11415040	767	EIRVINQTTTCENLLPQQTIPRMMCVGFLSGGVDSQCGDSGGPLSSVEADGRIFOAGVVS	826			
	gi 7363445	767	EIRVINQTTTCEDLMPQQTIPRMMCVGFLSGGVDSQCGDSGGPLSSAEKGRMFOAGVVS	826			
	gi 16758444	767	EIRVINQTTTCEELLPQQTIPRMMCVGFLSGGVDSQCGDSGGPLSSVEKGRIFOAGVVS	826			
60		850	860				
	NOV21a	228	GYGCGRPHFPGVYTRVAAVRGWIGQHIQE	256			
	NOV21b	232	GYGCGRPHFPGVYTRVAAVRGWIGQHIQE	260			
65	gi 12836503	772	GLCCGRPNFFGVYTRVTRVINWIOQVLT-	799			
	gi 10257390	827	GDGCAQRNKPFGVYTRIPLPFRDWIKENTGV	855			
	gi 11415040	827	GDGCAQRNKPFGVYTRIPLPFRDWIKENTGV	855			
	gi 7363445	827	GEGCAQRNKPFGVYTRIPLPFRDWIKEHTGV	855			
	gi 16758444	827	GEGCAQRNKPFGVYTRIPEVRDWIKEQTGV	855			

Tables 21G-H lists the domain descriptions from DOMAIN analysis results against NOV21. This indicates that the NOV21 sequence has properties similar to those of other proteins known to contain this domain.

Table 21G Domain Analysis of NOV21

gnl|Smart|smart00020, Tryp_SpC, Trypsin-like serine protease; Many of these are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. (SEQ ID NO:812)
CD-Length = 230 residues, 100.0% aligned
Score = 221 bits (563), Expect = 4e-59

5	NOV21:	27	RIVGGSAAGRGEWPQVSLWLRREHRCGAVLVAERWLLSAAHCFDVGDPKQWAAFLGT	86
	Sbjct:	1	RIVGGSEANIGSFPWQVSLQYRGGRHFCGGLISPRWVLTAAHCYV-GSAPSSIRVRLGS	59
10	NOV21:	87	PFL-SGAEGQLERVARIYKHPFYNYLTLDYDVALLELAGPVRRSRLVRPICLPEPAPRPP	145
	Sbjct:	60	HDLSSGEETQTVKVKVIVHPNPNSTYDNDIALLLKLEPVTLSDTVRLPICLPSSGYNVP	119
15	NOV21:	146	DGTRCVITGWGVSRE-GGSMARQLQKAAVRLLESEQTCRRFYPVQISSRISEPPFFSPQ--	202
	Sbjct:	120	AGTTCTVSGWGRTSESSGSLPDTLQEVNVPVSNATCRRAYSGGPAITDMLCAGGLEGG	179
20	NOV21:	203	----QGDAGGPLACREPSGRWVLTGVTSWG-YGCGRPHFPGVYTRVAAVRGWI	250
	Sbjct:	180	KDACQGDSGGPLVCN--DPRWVLVGIVSWGSGYGCARPKNKPGVYTRVSSYLDWI	230

Table 21H Domain Analysis of NOV21

gnl|Pfam|pfam00089, trypsin, Trypsin. Proteins recognized include all proteins in families S1, S2A, S2B, S2C, and S5 in the classification of peptidases. Also included are proteins that are clearly members, but that lack peptidase activity, such as haptoglobin and protein Z (PRTZ*). (SEQ ID NO:813)
CD-Length = 217 residues, 100.0% aligned
Score = 177 bits (448), Expect = 9e-46

25	NOV21:	28	IVGGSAAGRGEWPQVSLWLRREHRCGAVLVAERWLLSAAHCFDVGDPKQWAAFLGTP	87
	Sbjct:	1	IVGGREAQAGSFPWQVSLQVSSG-HFCGGLISENWVLTAAHCVS---GASSVRVVLGEH	56
30	NOV21:	88	FLSGAEGQLER--VARIYKHPFYNYLTLDYDVALLELAGPVRRSRLVRPICLPEPAPRPP	145
	Sbjct:	57	NLGTTEGTEQKFDVKIIVHPNPNPDT--NDIALLLKLSKSPVTLGDTVRLPICLPSSASDLP	114
35	NOV21:	146	DGTRCVITGWGVSREGGSMARQLQKAAVRLLESEQTCRRFYPVQISSR---ISEPPFFSPQ	202
	Sbjct:	115	VGTTCSVSGWGRTKNLGT-SDTLQEVVVPVVSRETCSRAYSAGGTVDTDMICAGALGGKDAC	173
	NOV21:	203	QGDAGGPLACREPSGRWVLTGVTSWGYGCGRPHFPGVYTRVAAVRGWI	250
	Sbjct:	174	QGDGGPLVCSGDG----ELVGIVSWGYGCAVGNYPGVYTRVSRYLWDWI	217

Proteolytic enzymes that exploit serine in their catalytic activity are ubiquitous, being found in viruses, bacteria and eukaryotes . They include a wide range of peptidase activity, including exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activity. Over 20 families (denoted S1 - S27) of serine protease have been identified, these being grouped into 6
5 clans on the basis of structural similarity and other functional evidence.

Trypsin is a tetrameric serine protease that is concentrated and stored selectively in the secretory granules of all types of mast cells, from which it is secreted during mast cell degranulation. Its exclusive presence in mast cells permits its use as a specific clinical indicator of mast cell activation by measurement of its level in biologic fluids and as a
10 selective marker of intact mast cells using immunohistochemical techniques with antitrypsin antibodies.

In addition, NOV21 nucleic acids and polypeptides are useful, inter alia, as novel members of the protein families according to the presence of domains and sequences related to previously described proteins. For example, NOV21 nucleic acids and polypeptides contain a
15 structural motif that is characteristic of protein sbelonging to the serine protease family of proteins. Accordingly, NOV21 may be useful in the same ways other members of this family are useful as detailed above.

The disclosed NOV21 nucleic acid of the invention encoding a Adrenal secretory serine protease -like protein includes the nucleic acid whose sequence is provided in Table
20 21A, 21C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 21A or 21C while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described,
25 including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
30 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV21 protein of the invention includes the Adrenal secretory serine protease -like protein whose sequence is provided in Table 21B or 24D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 21B or 21D while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 54 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Adrenal secretory serine protease -like protein (NOV21) is a member of a "Adrenal secretory serine protease family". Therefore, the NOV21 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV21 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, endometriosis, fertility, anemia, ataxia-telangiectasia, autoimmune disease, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphoedema, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, and/or other diseases and pathologies.

NOV21 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV21 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV21 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV22

NOV22 includes three novel adrenal secretory serine protease-like proteins disclosed below. The disclosed sequences have been named NOV22a, and NOV22b.

NOV22a

A disclosed NOV22a nucleic acid of 796 nucleotides (also referred to as CG56643-01) encoding a novel adrenal secretory serine protease-like protein is shown in Table 22A. An open reading frame was identified beginning with an ACC initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 763-765. The start and stop codons are shown in bold in Table 22A, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon of NOV22a is not a traditional initiation codon, NOV22a could be a partial reading frame that extends further in the 5' direction.

Table 22A. NOV22a nucleotide sequence (SEQ ID NO:91).

ACCCGAGCAGGCCAAGATCCCCAGACCTGGTCTGTGTCCTCCTCCAGAATGTGGGGCCAGGCCTGCAATG
GAGAAGCCCCACCCGGGTCGTGCGCGGGTTCGGAGCTGCCTCCGGGGAGGTGCCCTGGCAGGTCAGCCTGAAG
GAAGGGTCCCGGCACTTCTGCGGAGCAACTGTGGTGGGGACCGCTGGCTGCTGTCTGCCGCGCCACTGTCTC
CATAGCACGAAGGTGGAGCAGGTTCGGGCCCCACCTGGGCACTGCGTCCCTCCTGGGCTGGGCGGGAGCCCG
GTGAAGATCGGGCTGCGGCGGGTAGTGCTGCACCCCCCTCTACAACCTGGCATCCTGGACTTCGACCTGGCT
GTCCTGGAGCTGGCCAGCCCCCTGGCCTTCAACAAATACATCCAGCCTGTCTGCCTGCCCTGGCCATCCAG
AAGTTCCTGTGGGCGGGAAGTGCATGATCTCCGGATGGGGAATACGCAGGAAGGAAATCTGCAGAAGGCG
TCCGTGGGCATCATAGACCAGAAAACCTGTAGTGTGCTCTACAACCTTCTCCCTCACAGACCGCATGATCTGC
GCAGGCTTCTGGAAGGCAAAGTCGACTCCTGCCAGGGTGACTCTGGGGGCCCCCTGGCCTGCGAGGAGGCC
CCTGGCGTGTTTTATCTGGCAGGGATCGTGAGCTGGGGTATTTGGCTGCGCTCAGGTTAAGAAGCCGGGCGTG
TACACGCGCATCACCAGGCTAAAGGCTGGATCATCCAGGAGT**GACCACC**CGTGACTGCCAGGCCGAGAC
TCTA

In a search of public sequence databases, the NOV22a nucleic acid sequence, located on chromosome 19, has 278 of 428 bases (64%) identical to a gb:GENBANK-
ID:E13204|acc:E13204.1 mRNA from *Homo sapiens* (Human cDNA encoding a serine
protease) ($E = 1.6e^{-29}$).

The disclosed NOV22a polypeptide (SEQ ID NO:92) encoded by SEQ ID NO:91 has 254 amino acid residues and is presented in Table 22B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV22a has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.5090. Alternatively, NOV22a may also localize to the cytoplasm with a certainty of 0.4500, to the lysosome (lumen) with a certainty of 0.2082, or the mitochondrial matrix space with a certainty of 0.1000.

Table 22B. Encoded NOV22a protein sequence (SEQ ID NO:92).

TRAGQDPQTWSCVLLPECGARPAMEKPTRVVRGFGAASGEVPWQVSLKEGSRHFCGATVVGDRWLLSAAHCF
HSTKVEQVRAHLGTASLLGLGGSPVKIGLRRVVLHPLYNPGILDVFLAVLELASPLAFNKYIQPVCLPLAIQ

KFPVGRKCMISGWGNTQEGNLQKASVGIIDQKTCVLYNFSLTDRMICAGFLEGKVDSCQGDSSGGPLACEEA
PGVFYLAGIVSWGIGCAQVKKPGVYTRITRLKGWIIQE

A search of sequence databases reveals that the NOV22a amino acid sequence has 100 of 241 amino acid residues (41%) identical to, and 149 of 241 amino acid residues (61%) similar to, the 273 amino acid residue ptrn:TREMBLNEW-ACC:BAB20278 protein from *Mus musculus* (Mouse) (Type 1 Spinesin) ($E = 3.1e^{-49}$).

The adrenal secretory serine protease disclosed in this invention is predicted to be expressed in at least the following tissues: Ovary, kidney, breast, lung, muscle, liver, spleen, blood, lymphocyte. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV22b

In the present invention, the target sequence identified previously, NOV22a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV22b. This differs from the previously

identified sequence (NOV22a) in having 43 additional aminoacids and different N and C terminus.

A disclosed NOV22b nucleic acid of 992 nucleotides (also referred to as CG56643-02) encoding a novel adrenal secretory serine protease-like protein is shown in Table 22C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 101-103 and ending with a TAA codon at nucleotides 920-922. The start and stop codons are shown in bold in Table 22C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 22C. NOV22b nucleotide sequence (SEQ ID NO:93).

GCTAGTCTATCCCGAGACCCCTCCCACTCCAACAGTTAATGCTTCCCTTGACCTCAGAATGGCCTCTACAC
CTTACCCAGGTGCTAGGGCGGCAGCCCCATGGGGACAGTGGGGAGACTCTTGCGCTCTGAGCGGGCCATCAG
GCCCCACTCTCTCACTCTGTGGCTTTGTGAGATTCTGCAACTCTGTGAGCCCTGGTTTCTTCGTCTGTG
GGGTGGGGATGCTGCATCTCGGGGCTGTTATCGGAGCGGAAGTGGAGCTGCTCTGATGATCACTGTGCACGT
GGCCTTTCTGGCTCTTCCCTGGTAGCCACCAAGCCGAGCTCCTGCAGAAGGCGTCCGTGGGCATCATAGA
CCAGAAAACCTGTAGTGTGCTCTACAACCTTCTCCCTCAGACCCGATGATCTGCGCAGGCTTCTGGAAGG
CAAAGTCGACTCCTGCCAGGGTGACTCTGGGGGCCCTGGCCTGCGAGGAGGCCCTGGCGTGTCTTATCT
GGCAGGGATCGTGAGCTGGGGTATTGGCTGCGCTCAGGTTAAGAAGCCGGGCGTGACACGCGCATCACCAG
GCTAAAGGGCTGGATCCTGGAGATCATGTCCTCCAGCCCTTCCCATGTCTCCCCCTCGACCAAGGAT
GCTGGCCACCACAGCCCCAGGACGACAGCTGGCCTCACAGTCCCGGGGGCCACACCCAGCAGACCCACCCC
TGGGGCTGCCAGCAGGTGACGGGCCAACCTGCCAACTCAACCTTATCTGCCGTGAGCACCCTGCTAGGGG
ACAGACGCCATTTCCAGACGCCCGGAGGCCACACACACCCAGCTACCAGGTACCGGGAGAGACGGAGG
GATCCCTGGGAGTGGAGGGTCCCATGTTAATCAGCCTGGGCTGCCTAACAAGACATAACGTCGTCCACTTGG
GGAGGCCGAGGCGGCGGATCAAGAGGTGAGGATCGAGACCATCCTGGCGAACA

In a search of public sequence databases, the NOV22b nucleic acid sequence, located on chromosome 19, has 203 of 294 bases (69%) identical to a gb:GENBANK-ID:AF133086|acc:AF133086.1 mRNA from *Homo sapiens* (membrane-type serine protease 1 mRNA, complete cds) ($E = 3.6e^{-16}$).

The disclosed NOV22b polypeptide (SEQ ID NO:94) encoded by SEQ ID NO:93 has 273 amino acid residues and is presented in Table 22D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV22b has A signal peptide and is likely to be localized to the mitochondrial inner membrane with a certainty of 0.8723. Alternatively, NOV22b may also localize to the plasma membrane with a certainty of 0.6500, to the mitochondrial intermembrane space with a certainty of 0.5053, or the mitochondrial matrix space with a certainty of 0.3617. The most likely cleavage site for NOV22b is between positions 43 and 44: GDA-AS.

Table 22D. Encoded NOV22b protein sequence (SEQ ID NO:94).

MGTVGRLLRSEAIRPTSSSLCGFVRFLQLCEPWFRLRLWGGDAASRGCYRSGTGAALMITVHVAFALSLVA
TKPELLQKASVGIIDQKTCVLYNFSLTDRMICAGFLEGKVDSCQDSSGGLACEEAPGVFYLAGIVSWGIG
CAQVKKPGVYTRITRLKGWILEIMSSQPLPMSPPSTTRMLATTSPTTAGLTVPGATPSRPTPGAASRVTGQ
PANSTLSAVSTTARGQTPFPDAPEATHTTQLPGTGRDGGIPGSGSHVNQPLPNKT

A search of sequence databases reveals that the NOV22b amino acid sequence has 49 of 90 amino acid residues (54%) identical to, and 63 of 90 amino acid residues (70%) similar to, the 277 amino acid residue ptnr:SPTREMBL-ACC:O96899 protein from *Scolopendra subspinipes* (Plasminogen Activator Spa) ($E = 4.3e^{-24}$).

NOV22b is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

NOV22c

A disclosed NOV22c nucleic acid of 912 nucleotides (also referred to as CG56643-03) encoding a novel adrenal secretory serine protease-like protein is shown in Table 22E. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 77-79 and ending with a TAA codon at nucleotides 896-898. The start and stop codons are shown in bold in Table 22E, and the 5' and 3' untranslated regions, if any, are underlined.

Table 22E. NOV22c nucleotide sequence (SEQ ID NO:95).

CACTCCAACAGTTAATGCTTCCCTTGACCTCAGAATGGCCTCCTACACCTTACCCAGGTGCTAGGGCGGCAG
CCCCATGGGGACAGTGGGGAGACTCTTGCGCTCTGAGCGGGCCATCAGGCCACCTCCTCCTCACTCTGTGG
CTTTGTGAGATTCTTGCAACTCTGTGAGCCCTGGTTTCTTCGTCTGTGGGGTGGGGATGCTGCATCTCGGGG
CTGTTATCGGAGCGGAAGTGGAGCTGCTCTGATGATCACTGTGCACGTGGCCTTTCTGGCTCTTTCCCTGGT
AGCCACCAAGCCCGAGCTCCTGCAGAAGGCGTCCGTGGGCATCATAGACCAGAAAACCTGTAGTGTGCTCTA
CAACTTCTCCCTCACAGACCGCATGATCTGCGCAGGCTTCTTGGGAAGGCAAAGTCGACTCCTGCCAGGGTGA
CTTGGGGGGCCCCCTGGCCTGCGAGGAGGCCCTGGCGTGTTTTATCTGGCAGGGATCGTGAGCTGGGGTAT
TGGCTGCGCTCAGGTAAAGAAGCCGGGCGTGTACACGCGCATCACCAGGCTAAAGGGCTGGATCCTGGAGAT
CATGTCCTCCCAGCCCCCTTCCCATGTCTCCCCCTCGACCACAAGGATGCTGGCCACCACCAGCCCCAGGAC
GACAGCTGGCCTCACAGTCCCGGGGGCCACACCCAGCAGACCCACCCCTGGGGCTGCCAGCAGGGTGACGGG
CCAACCTGCCAACTCAACCTTATCTGCCGTGAGCACCCTGCTAGGGGACAGACGCCATTTCAGACGCCCC
GGAGGCCACCACACACACCCAGCTACCAGGTACCGGGAGAGACGGAGGGATCCCTGGGAGTGAGGGTCCCA
TGTTAATCAGCCTGGGCTGCCAACAAGACATA**AC**CTCGTCCACTTTG

In a search of public sequence databases, the NOV22c nucleic acid sequence, located on chromosome 19, has 203 of 294 bases (69%) identical to a gb:GENBANK-ID:E13204|acc:E13204.1 mRNA from *Homo sapiens* (Human cDNA encoding a serine protease) ($E = 1.3e^{-18}$).

The disclosed NOV22c polypeptide (SEQ ID NO:96) encoded by SEQ ID NO:95 has 273 amino acid residues and is presented in Table 22F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV22c has A signal peptide and is

likely to be localized to the mitochondrial inner membrane with a certainty of 0.8723. Alternatively, NOV22c may also localize to the plasma membrane with a certainty of 0.6500, to the mitochondrial intermembrane space with a certainty of 0.5053, or the mitochondrial matrix space with a certainty of 0.3617. The most likely cleavage site for NOV22c is between positions 43 and 44: GDA-AS.

Table 22F. Encoded NOV22c protein sequence (SEQ ID NO:96).

MGTVGRLLRSEAIRPTSSSLCGFVRFLQLCEPWFLRLWGGDAASRGCYRSGTGAALMITVHVAFLALSLVA
TKPELLQKASVGIIDQKTCVLYNFSLTDRMICAGFLEGKVDSCQGDSSGGPLACEEAPGVFYLAGIVSWGIG
CAQVKKPGVYTRITRLKGWILEIMSSQPLPMSPSTTRMLATTSPRTTAGLTVPGATPSRPTPGAASRVGTQ
PANSTLSAVSTTARGQTPFPDAPEATHTHTQLPGTGRDGGIPSGGSHVNQPLPNKT

A search of sequence databases reveals that the NOV22c amino acid sequence has 49 of 90 amino acid residues (54%) identical to, and 63 of 90 amino acid residues (70%) similar to, the 277 amino acid residue ptnr:SPTREMBL-ACC:O96899 protein from *Scolopendra subspinipes* (Plasminogen Activator SPA) ($E = 4.5e^{-24}$).

NOV22c is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. .

The disclosed NOV22a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 22G..

Table 22G. BLAST results for NOV22a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16758444 ref NP_446087.1 (NM_053635)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [Rattus norvegicus]	855	109/251 (43%)	148/251 (58%)	7e-55
gi 7363445 ref NP_035306.2 (NM_011176)	protease, serine, 14 (epithin) [Mus musculus]	855	110/248 (44%)	150/248 (60%)	7e-54
gi 9757702 dbj BAB08218.1 (AB038498)	homolog of human MT-SP1 [Xenopus laevis]	845	113/261 (43%)	156/261 (59%)	2e-52
gi 10257390 gb AAG15395.1 AF057145_1 (AF057145)	serine protease TADG15 [Homo sapiens]	855	107/248 (43%)	145/248 (58%)	3e-52

gi 11415040 ref NP_068813.1 (NM_021978)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [<i>Homo sapiens</i>]	855	107/248 (43%)	145/248 (58%)	3e-52
---	---	-----	------------------	------------------	-------

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 22H. In the ClustalW alignment of the NOV22 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 22H. ClustalW Analysis of NOV22

- 1) Novel NOV22a (SEQ ID NO:92)
- 2) Novel NOV22b (SEQ ID NO:94)
- 3) Novel NOV22c (SEQ ID NO:96)
- 4) gi|16758444|ref|NP_446087.1| (NM_053635) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [*Rattus norvegicus*] (SEQ ID NO:414)
- 5) gi|7363445|ref|NP_035306.2| (NM_011176) protease, serine, 14 (epithin) [*Mus musculus*] (SEQ ID NO:413)
- 6) gi|9757702|dbj|BAB08218.1| (AB038498) homolog of human MT-SP1 [*Xenopus laevis*] (SEQ ID NO:415)
- 7) gi|10257390|gb|AAG15395.1|AF057145_1 (AF057145) serine protease TADG15 [*Homo sapiens*] (SEQ ID NO:411)
- 8) gi|11415040|ref|NP_068813.1| (NM_021978) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [*Homo sapiens*] (SEQ ID NO:412)

		10	20	30	40	50	60	
NOV22a							1
NOV22b	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----							1
NOV22c	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----							1
gi 16758444 ref	MGNNRGRKAGGGSODFCAGLKYNSRLNNMGFEEGVEFLPVNNAKOVEKRGERRWVVMVA	60						
gi 7363445 ref	MGSNRGRKAGGGSODFCAGLKYNSRLNNMGFEEGVEFLPVNNAKVEKRGERRWVVLVA	60						
gi 9757702 dbj	-----MKDSDSMMKYNNRPOSINGFEEGVEFLPATNSKKVEKTGPKKKLAFG	48						
gi 10257390 gb	MGSDRARKGGGPKDFCAGLKYNSRHEKVNGLLEEGVEFLPVNNVKKVEKHGPGRRWVVLAA	60						
gi 11415040 ref	MGSDRARKGGGPKDFCAGLKYNSRHEKVNGLLEEGVEFLPVNNVKKVEKHGPGRRWVVLAA	60						
		70	80	90	100	110	120	
NOV22a							1
NOV22b	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----							1
NOV22c	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----							1
gi 16758444 ref	VMFSFLLLSLMAGLLVWHFHYRNVRTOKVFNCHLRITNENFLDAYENSTSTEFISLASOV	120						
gi 7363445 ref	VLFSFLLLSLMAGLLVWHFHYRNVRVOKVFNCHLRITNETFLDAYENSTSTEFISLASOV	120						
gi 9757702 dbj	LVIGAALLSLTIGLLVWHFAYRNKPVNKLYTCYLTITANTPFIDAYENSTTAEFSDLSAKV	108						
gi 10257390 gb	VLIGLLLVLLGIGFLVWHLQYRDVVRVOKVFNCHLRITNENFVDAYENSTSTEFVSLASKV	120						
gi 11415040 ref	VLIGLLLVLLGIGFLVWHLQYRDVVRVOKVFNCHLRITNENFVDAYENSTSTEFVSLASKV	120						

		130	140	150	160	170	180	
5	NOV22a	1					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	1					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 16758444 ref	KEALKLMYSEVPVLGPGYHKKSTVTFSEGG---SVIAYYWSEFSIPPHLEEEVDRAVAVER	177					
	gi 7363445 ref	KEALKLLYNEVPVLGPGYHKKSAVTFSEGG---SVIAYYWSEFSIPPHLAEVDRAVAVER	177					
10	gi 9757702 dbj	IDTLQTVYNGNKDIAPYLQKCSISAFSEGGGNNVIGYYWSEFSVPAFREAFAFEKATSELK	168					
	gi 10257390 gb	KDALKLLYSGVPFLGPGYHKEAVTFSEGG---SVIAYYWSEFSIPQHLVEEAERVMAEER	177					
	gi 11415040 ref	KDALKLLYSGVPFLGPGYHKEAVTFSEGG---SVIAYYWSEFSIPQHLVEEAERVMAEER	177					
		190	200	210	220	230	240	
15	NOV22a	1					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	1					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 16758444 ref	VVTLPFRARALKSFVLTSVVAFFPIDPRMLQRTQDNSSCFALHARGRTVTRFTTTCGFPNPS	237					
	gi 7363445 ref	VVTLPFRARALKSFVLTSVVAFFPIDPRMLQRTQDNSSCFALHAHGAAVTRFTTTCGFPNPS	237					
20	gi 9757702 dbj	LPSVNPROR---TFALDSLVAYPTDQIARVFKNSSCAYFLHSSNGGVAKFSSGFPDPS	225					
	gi 10257390 gb	VVMLPFRARSLKSFVVTSSVAFPTDSKTVQRTQDNSSCFGLHARGVELMRFTTTCGFPDPS	237					
	gi 11415040 ref	VVMLPFRARSLKSFVVTSSVAFPTDSKTVQRTQDNSSCFGLHARGVELMRFTTTCGFPDPS	237					
25	NOV22a	1					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	1					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	1					
	gi 16758444 ref	YPAHARCQWVLRGDADSVLSLTFRSFDVAPCDGHSDSLVTIVYDSLSPMEFHAVVRLCGTF	297					
30	gi 7363445 ref	YPAHARCQWVLRGDADSVLSLTFRSFDVAPCDEHGSDDLVTIVYDSLSPMEFHAVVRLCGTF	297					
	gi 9757702 dbj	YPRNARCLWTLRADAGRIIHLHFKTKMEKCKPNGGDFVMVYDSLSPMEFRAOIRLCGTY	285					
	gi 10257390 gb	YPAHARCQWALRGDADSVLSLTFRSFDIASCDERGSDDLVTIVYNTLSPMEFHALVOLCGTY	297					
	gi 11415040 ref	YPAHARCQWALRGDADSVLSLTFRSFDIASCDERGSDDLVTIVYNTLSPMEFHALVOLCGTY	297					
35	NOV22a	1					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	15					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	15					
40	gi 16758444 ref	SPSYNLTFLLSSQNVFLVTLITNTDRRHHPGFEATTFQOLPKMSSCGGLLSEACGTFSSPYYP	357					
	gi 7363445 ref	SPSYNLTFLLSSQNVFLVTLITNTDRRHHPGFEATTFQOLPKMSSCGGFLSDTCGTFSSPYYP	357					
	gi 9757702 dbj	PPSYNLTFHSSQNVLLITLITNTERRHHPGFEATTFQOLPRMSSCGGRLRKAQCTFNSPYYP	345					
	gi 10257390 gb	PPSYNLTFHSSQNVLLITLITNTERRHHPGFEATTFQOLPRMSSCGGRLRKAQCTFNSPYYP	357					
	gi 11415040 ref	PPSYNLTFHSSQNVLLITLITNTERRHHPGFEATTFQOLPRMSSCGGRLRKAQCTFNSPYYP	357					
45	NOV22a	1					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	40					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	40					
50	gi 16758444 ref	GHYPPNINCTWNKVPNNRNKVKRFKLFYLVDPNIEVGSCTKDYVEINGEKFCGERSQFV	417					
	gi 7363445 ref	GHYPPNINCTWNKVPNNRNKVKRFKLFYLVDPNIEVGSCTKDYVEINGEKFCGERSQFV	417					
	gi 9757702 dbj	AHYPPSTESIWDIQVPDNKFKVKRFNMFYLAEPGVETKCTKDFVEIKGCKYCGEKEFFV	405					
	gi 10257390 gb	GHYPPNIDCTWNIEVPNNQHVKSFKFFYLLEPGVBPAGTCPKDYVEINGEKFCGERSQFV	417					
55	gi 11415040 ref	GHYPPNIDCTWNIEVPNNQHVKSFKFFYLLEPGVBPAGTCPKDYVEINGEKFCGERSQFV	417					
		430	440	450	460	470	480	
60	NOV22a	4					
	NOV22b	----- ----- ----- ----- ----- ----- ----- -----	54					
	NOV22c	----- ----- ----- ----- ----- ----- ----- -----	54					
	gi 16758444 ref	VSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRKDLRCDGWADC	477					
	gi 7363445 ref	VSSNSSKITVHFHSDHSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRKDLRCDGWADC	477					
65	gi 9757702 dbj	VSSNSSKMSVRFVSDQSYTDTGFTAELYSYEPNPNCPDQFTCRSGRCIRLDQKCDGWADC	465					
	gi 10257390 gb	VTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKDLRCDGWADC	477					
	gi 11415040 ref	VTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKDLRCDGWADC	477					
70	NOV22a	19					
		490	500	510	520	530	540	
		QD	PQW	SCVL	PECG			

NOV22b
 NOV22c
 gi | 16758444 | ref | PDYSDERHCRNATHQFMCKNQFCKPLFWVCDVNDGCGDSDEEGGCSCPAGSEFKCSNGKC 537
 gi | 7363445 | ref | PDYSDERYCRNATHQFTCKNQFCKPLFWVCDVNDGCGDSDEEGGCSCPAGSEFKCSNGKC 537
 gi | 9757702 | dbj | EDFSDEMSCTCTALQFRVNSKILCKPSYFELCDGVNDGCGDSDELACKCPNNFTFKCSNGKC 525
 gi | 10257390 | gb | TDHSDDELNCSCDAGHQFTCKNKFCCKPLFWVCDVNDGCGDNSDEQGCSCPAQTFRCNSNGKC 537
 gi | 11415040 | ref | TDHSDDELNCSCDAGHQFTCKNKFCCKPLFWVCDVNDGCGDNSDEQGCSCPAQTFRCNSNGKC 537

550 560 570 580 590 600
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | LPQSQQCNKGKDDCGDGSDEASCDNVNAVSTCKYTYRCNGLCLNKNPECDGKKDCSDGS 597
 gi | 7363445 | ref | LPQSOKCNKGKDDCGDGSDEASCDNVNVVSTCKYTYRCNGLCLSKNPECDGKTDSDGS 597
 gi | 9757702 | dbj | LPDSQKCDRVNDCGDGSDEAECDQVLTACTEYTYCKKNQCTTKNPECDGENDCSGDS 585
 gi | 10257390 | gb | LSKSQCCNKGKDDCGDGSDEASCPKVNVTCTKHTYRCNGLCLSKNPECDGKEDCSGDS 597
 gi | 11415040 | ref | LSKSQCCNKGKDDCGDGSDEASCPKVNVTCTKHTYRCNGLCLSKNPECDGKEDCSGDS 597

610 620 630 640 650 660
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | DE--KNCDCLRSFTQARVVGCTDADEGEWPQVSVLHLAGQGHICGASLISPDWLVSAA 655
 gi | 7363445 | ref | DE--KNCDCLRSFTQARVVGCTDADEGEWPQVSVLHLAGQGHICGASLISPDWLVSAA 655
 gi | 9757702 | dbj | DENAAKCNCGKRPFTKKSRIIVGGVNDTGEFPQVSVLHAKGNKHTCGASLGFTMLISAA 645
 gi | 10257390 | gb | DE--KDCCCLRSFTQARVVGCTDADEGEWPQVSVLHLAGQGHICGASLISPDWLVSAA 655
 gi | 11415040 | ref | DE--KDCCCLRSFTQARVVGCTDADEGEWPQVSVLHLAGQGHICGASLISPDWLVSAA 655

670 680 690 700 710 720
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | HCFQDETIFKYSDDHTMTWAFGLGLDQSKRSASGVQEHKLRILITHEFNDFTFDYDIAL 715
 gi | 7363445 | ref | HCFQDDKNFKYSDDHTMTWAFGLGLDQSKRSASGVQEHKLRILITHEFNDFTFDYDIAL 715
 gi | 9757702 | dbj | HCFQDDHQMRYSDASLWTAYLGLHDAQNLTKDVVERRTKRIMAHIGFNDNTVDNDIAL 705
 gi | 10257390 | gb | HCYIDDRGFRYSDDPTQWTAFLGLHDAQSQRAPGVQERRLKRIISHFFNDFTFDYDIAL 715
 gi | 11415040 | ref | HCYIDDRGFRYSDDPTQWTAFLGLHDAQSQRAPGVQERRLKRIISHFFNDFTFDYDIAL 715

730 740 750 760 770 780
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | ELEKPAEYSTVVRP---ICLPDATHVFPAGKAIWVTGWGHTKEGGTGALILQKEIRVIN 772
 gi | 7363445 | ref | ELEKSVEYSTVVRP---ICLPDATHVFPAGKAIWVTGWGHTKEGGTGALILQKEIRVIN 772
 gi | 9757702 | dbj | ELEKPAEYNDFIQF---VCIPESTHDFPVGKPIWVTGWGALKEGGAIVILQKAEIRIIN 762
 gi | 10257390 | gb | ELEKPAEYSSMVRP---ICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVIN 772
 gi | 11415040 | ref | ELEKPAEYSSMVRP---ICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVIN 772

790 800 810 820 830 840
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | QTTCEELLPQOITPRMNCVGFSLGGVDSQCGDSGGPLSSVEADGRIFQAGVVSWEDEGCA 832
 gi | 7363445 | ref | QTTCEDLMPQOITPRMNCVGFSLGGVDSQCGDSGGPLSSAEKGRMFQAGVVSWEDEGCA 832
 gi | 9757702 | dbj | QTECNKLLDGLTTPRMLCAGFVSGGIDACQDSSGGPLSSVELNNKVYLAGVVSWEDEGCA 822
 gi | 10257390 | gb | QTTCEENLLPQOITPRMNCVGFSLGGVDSQCGDSGGPLSSVEADGRIFQAGVVSWEDEGCA 832
 gi | 11415040 | ref | QTTCEENLLPQOITPRMNCVGFSLGGVDSQCGDSGGPLSSVEADGRIFQAGVVSWEDEGCA 832

850 860
 NOV22a
 NOV22b
 NOV22c
 gi | 16758444 | ref | RNKPGVYTRIPVVRDWIKEQTGV 855
 gi | 7363445 | ref | RNKPGVYTRIPVVRDWIKEQTGV 855

```

gi|9757702|dbj| RNKPGVYTKVSMRDWSKDKTGH 845
gi|10257390|gb| RNKPGVYTRIPLERDWIKENTGV 855
gi|11415040|ref| RNKPGVYTRIPLERDWIKENTGV 855

```

5

Tables 22I-J lists the domain descriptions from DOMAIN analysis results against NOV22. This indicates that the NOV22 sequence has properties similar to those of other proteins known to contain this domain.

Table 22I Domain Analysis of NOV22

gnl|Smart|smart00020, Tryp_SPC, Trypsin-like serine protease; Many of these are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. (SEQ ID NO:812)
 CD-Length = 230 residues, 100.0% aligned
 Score = 220 bits (560), Expect = 9e-59

10

```

NOV22: 29  RVRGFGAASGEVPWQVSLK-EGSRHFCGATVVGDRWLLSAAHCFHSTKVEQVRAHLGTA 87
          | + | | | | | | | + | | | | | + + + | + | + | | | + + + | | +
Sbjct: 1   RIVGGSEANIGSFPWQVSLQYRGGRHFCGGLISPRWVLTAACHCVGSAPSSIRVRLGSH 60

15 NOV22: 88  SLLGLGGSPVKIGLRRVVLHPLYNPGILDLDLAVLELASPLAFNKYIQPVCLPLAIQKFP 147
          | + + + | + + | | | | | | + | + | + | + | + + | + | | | + |
Sbjct: 61  DLSSGEE-TQTVKVSQVIVHPNPNSTYDNDIALLLKLEPVTLSDTVVRPICLPSSGYNVP 119

20 NOV22: 148 VGRKCMISGWGNTQEGN-----LQKASVGIIDQKTCVLY--NFSLTDRMICAGFLEGG 199
          | | + | | | | | + | | + + | + | | | | | | + + | | + | | | | |
Sbjct: 120 AGTTCTVSGWGRTESSGSLPDTLQEVNVPVIVSNATCRRAYSGGPAITDNMLCAGGLEGG 179

NOV22: 200  VDSCQGDSSGGLACEEAPGVFYLAGIVSWG-IGCAQVKKPGVYTRITRLKGWI 251
          | + | | | | | | | | + + | | | | | | | | + | | | | | + + | |
25 Sbjct: 180 KDACQGDSSGGLVCND--PRWVLVGIVSWGSGCARPNKPGVYTRVSSYLDWI 230

```

Table 22J Domain Analysis of NOV22

gnl|Pfam|pfam00089, trypsin, Trypsin. Proteins recognized include all proteins in families S1, S2A, S2B, S2C, and S5 in the classification of peptidases. Also included are proteins that are clearly members, but that lack peptidase activity, such as haptoglobin and protein Z (PRTZ*). (SEQ ID NO:813)
 CD-Length = 217 residues, 100.0% aligned
 Score = 192 bits (488), Expect = 2e-50

```

NOV22: 30  VVRGFGAASGEVPWQVSLKEGSRHFCGATVVGDRWLLSAAHCFHSTKVEQVRAHLGTASL 89
          + | | | + | | | | | + | | | | + + + + | + | + | | | + | | + |
30 Sbjct: 1   IVGGREAQAGSFPWQVSLQVSSGHFCGGLISENWWVLTAACHCVSGASSVRVVL--GEHNL 58

NOV22: 90  LGLGGSPVKIGLRRVVLHPLYNPGILDLDLAVLELASPLAFNKYIQPVCLPLAIQKFPVG 149
          | + | + + + + + | | | | | + | + | + | | + + + | + | | | | |
Sbjct: 59  GTTEGTEQKFDVKKIIVHPNPNPD--TNDIALLLKLSPTLGDTVRPICLPSASSDLFVG 116

35 NOV22: 150 RKCMIISGWGNTQEGN----LQKASVGIIDQKTCVLYNFSLTDRMICAGFLEGGKVDSCQG 205
          | + | | | | + | | + | | + + | | | + + | | | | | | | | + | | |
Sbjct: 117 TTCVSVSGWGRTKNLGTSDTLQEVVVPIVSRETCSRAYSAGGTVTDTMICAGALGGK-DACQG 175

```

NOV22: 206 DSGGPLACEEAPGVFYLAGIVSWGIGCAQVKKPGVYTRITRLKGWI 251
 ||||| | + | ||||| || | |||||++| |
 Sbjct: 176 DSGGPLVCS DG---ELVGIVSWGYGCAVGNYPGVYTRVSRYL DWI 217

5 Proteolytic enzymes that exploit serine in their catalytic activity are ubiquitous, being found in viruses, bacteria and eukaryotes [1]. They include a wide range of peptidase activity, including exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activity. Over 20 families (denoted S1 - S27) of serine protease have been identified, these being grouped into 6 clans on the basis of structural similarity and other functional evidence [1].

10 Trypsin is a tetrameric serine protease that is concentrated and stored selectively in the secretory granules of all types of mast cells, from which it is secreted during mast cell degranulation. Its exclusive presence in mast cells permits its use as a specific clinical indicator of mast cell activation by measurement of its level in biologic fluids and as a selective marker of intact mast cells using immunohistochemical techniques with antitrypsin antibodies. Vanderslice [2] demonstrated the existence of multiple trypsins. In this respect, mast cell trypsin is like other serine proteases such as glandular kallikrein and trypsin, which are also members of multigene families. Miller et al. [3] mapped both alpha-trypsin and beta-trypsin to human chromosome 16 by PCR analysis of DNA from human/hamster somatic cell hybrids. Miller et al. [3] cloned a second cDNA for human trypsin, called beta-trypsin, from 15 a mast cell cDNA library. The 1,142 bases of beta-trypsin were found to encode a 30-amino acid leader sequence of 3,089 daltons and a 245-amino acid catalytic region of 27,458 daltons. The amino acid sequence of beta-trypsin was found to be 90% identical with that of alpha-trypsin, the first 20 amino acids of the catalytic portions being 100% identical. Both alpha- and beta-trypsin sequences were localized to human chromosome 16 by analysis of DNA 20 preparations from 25 human/hamster somatic cell hybrids by PCR.

25 Because of the presence of the trypsin domains and the homology to the adrenal secretory serine protease, we anticipate that the novel sequence described here will have useful properties and functions similar to these genes.

The disclosed NOV22 nucleic acid of the invention encoding a Adrenal secretory 30 serine protease -like protein includes the nucleic acid whose sequence is provided in Table 22A, 25C, 25E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 22A, 25C, or 25E while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a fragment of such a nucleic acid. The 35 invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids

just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
5 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV22 protein of the invention includes the Adrenal secretory serine
10 protease -like protein whose sequence is provided in Table 22B, 25D, or 25F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 22B, 25D, or 25F while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 57 percent of the
15 residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Adrenal secretory serine protease -like protein (NOV22) is a member of a "Adrenal secretory serine protease family". Therefore,
20 the NOV22 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene
25 delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV22 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, endometriosis, fertility, anemia, ataxia-telangiectasia, autoimmune disease,
30 hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphoedema, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, and/or other diseases and pathologies.

NOV22 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV22 substances for use in

therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV22 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV23

NOV23 includes three novel serine protease DESC1 protease-like proteins disclosed below. The disclosed sequences have been named NOV23a, NOV23b, NOV23c, and NOV23d.

NOV23a

The disclosed NOV23a nucleic acid of 1546 nucleotides (also referred to as CG56647-02) encoding a novel serine protease DESC1-like protein is shown in Table 23A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 101-103 and ending with a TAG codon at nucleotides 1481-1483. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 23A, and the start and stop codons are in bold letters.

Table 23A. NOV23a Nucleotide Sequence (SEQ ID NO:97)

GCCCCTGCCATAGGAGGCGGGGACTGTCAATTTACCGTCTCCTGATGCCATTCCAGAGGTTACGCCCTGA
 AGTCAGCTCAGATCCTGGGCCAGGCACTGCATGGGAGACAGGCATGAGCAGGACCTCTTTCTGCCTTCGA
 GGAAACACGGGGGCATCTGGGGCTCACTTGGCACTCATCCACCTTGTGCTGTACCTGGGGACCTCCGGC
 CTTCCTCTCTACACAGGGCTTCCACGTGGACCACACGGCCGAGCTGCGGGGAATCCGGTGGACCAGCAGT
 TTGCGGCGGGAGACCTCGGACTATCACGCGACGCTGACGCCACCCTGGAGGCACTGTTTGTAACTAGTT
 TTCAGAAGACAGAGTTAGAGGCAAGCTGCGTGGGTGCTCGGTACTGAATTATAGGGATGGGAATCCAG
 TGTCTCGTACATTTCCAGCTGCACTTCTGCTGCGACCCCTCCAGACGCTGAGCCTGGGCCTGGAGGAG
 GAGCTATTGCAGCGAGGGATCCGGGCAAGGCTGCGGGAGCACGGCATCTCCCTGGCTGCCTATGGCACA
 TTGTGTCGGCTGAGCTCACAGGTAGACATAAGGGACCCCTTGGCAGAAAGAGACTTCAAATCAGGTCGCTG
 TCCAGGGAACCTCTTTCTGCGGGAACAGCCAGTGTGTGACCAAGGTGAACCCGGAGTGTGACGACCAG
 GAGGACTGCTCCGATGGGTCCGACGAGGCGCACTGCGAGTGTGGCTTGCAGCCTGCCTGGAGGATGGCCG
 GCAGGATCGTGGGCGGCATGGAAGCATCCCCGGGGGAGTTTCCGTGGCAAGCCAGCCTTCGAGAGAACA
 GGAGCACTTCTGTGGGGCCGCCATCATCAACGCCAGGTGGCTGGTGTCTGCTGCTCACTGCTTCAATGAG
 TTCCAAGACCCGACGAAGTGGGTGGCTTACGTGGGTGCGACCTACCTCAGCGGCTCGGAGGCCAGCACCG
 TGCGGGCCAGGTGGTCCAGATCGTCAAGCACCCCTGTACAACGCGGACACGGCCGACTTTGACGTGGC
 TGTGCTGGAGCTGACAGCCCTCTGCCTTTTCGGCCGGCACATCCAGCCGCTGTGCTGCCGCTGCCACA
 CACATCTTCCCACCCAGCAAGAAGTGCTGATCTCAGGCTGGGGCTACCTCAAGGAGGACTTCGTGGTCA
 AGCCAGAGGTGTGTCAGAAAGCCACTGTGGAGCTGCTGGACCAGGCACTGTGTGCCAGCTTGTACGGCCA
 TTCACTCACTGACAGGATGGTGTGCGCTGGCTACCTGGACGGGAAGGTGGACTCCTGCCAGGGTGACTCA
 GGAGGACCCCTGGTCTGCGAGGAGCCCTCTGGCCGGTTCTTTCTGGCTGGCATCGTAGCTGGGGAATCG
 GGTGTGCGGAAGCCCGGCTCCAGGGGTCTATGCCGAGTACCAGGCTACGTGACTGGATCCTGGAGGC
 CACCGAAAGGTAGAAAGATGATGTACGTGCCTATCTTGATTTAGGGAGAACGGATATCGTCATAGTATCTT
 CATAAT

The disclosed NOV23a nucleic acid sequence, located on chromosome 19, has 356 of 566 bases (62%) identical to a gb:GENBANK-ID:AF133086|acc:AF133086.1 mRNA from *Homo sapiens* (membrane-type serine protease 1 mRNA, complete cds) ($E = 1.1e^{-23}$).

A disclosed NOV23a polypeptide (SEQ ID NO:98) encoded by SEQ ID NO:97 is 460 amino acid residues and is presented using the one-letter amino acid code in Table 23B. Signal P, Psort and/or Hydropathy results predict that NOV23a contains no signal peptide and is likely to be localized in the microbody (peroxisome) with a certainty of 0.5387.

Table 23B. Encoded NOV23a protein sequence (SEQ ID NO:98).

<p>MGDRHEQDLFLPSRKTRGHLGLTWHSSSTLCCTWGPPAFLSTQGFHVDHTAELRGIRWTSSLRRETSYHRTLTPT LEALFVSSFQKTELEASCVGCSVLNYRDGNSSVLVHFQHLFLRLPLQTLGLGLEEELLQRGIRARLREHGISLAA YGTIVSAELTGRHKGPLAERDFKSGRCPGNSFSCGNSQCVTKVNPECDDQEDCSGSDGSEAHCECGLQPAWRMAGR IVGMEASPGFEPWQASLRENKEHFCAAIINARWLVSAAHCFNEFDPTKWVAYVGATYLSGSEASTVRAQVVQ IVKHPLYNADTADFDVAVLELTSPLPFGRHIQPVCLPAATHIFPPSKKCLISGWGYLKEDFVVKPEVLQKATVEL LDQALCASLYGHSLTDRMVCAGYLDGKVDSCQGDGGGGLVCEEPSGRFFLAGIVSWGIGCAEARPGVYARVTRL RDWILEATER</p>
--

The disclosed NOV23a amino acid sequence has 112 of 248 amino acid residues (45%) identical to, and 157 of 248 amino acid residues (63%) similar to, the 422 amino acid residue ptnr:SPTREMBL-ACC:Q9UL52 protein from *Homo sapiens* (Human) (serine protease DESC1) ($E = 1.1e^{-58}$).

NOV23a is predicted to be expressed in at least Ovary, kidney, breast, lung, muscle, liver, spleen, blood and lymphocyte. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

NOV23b

A disclosed NOV23b nucleic acid of 1777 nucleotides (also referred to as CG56647-03) encoding a novel serine protease DESC1-like protein is shown in Table 23C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 101-103 and ending with a TAG codon at nucleotides 1631-1633. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 23C, and the start and stop codons are in bold letters.

Table 23C. NOV23b Nucleotide Sequence (SEQ ID NO:99)

<p> <u>GGCCCTGCCATAGGAGGCGGGGACTGTCATTTACCGTCTCCTGATGCCATTCCAGAGGTTACGCCCTGA</u> AGTCAGCTCAGATCCTGGGCCAGGCAGCTGCATGGGAGACAGGCATGAGCAGGACCTCTTCTGCCTTCGA GGAAACACGGGGGCATCTGGGGCTCACTTGGCACTCATCCACCTTGTGCTGTACCTGGGGACCTCCGGC CTTCCTCTACACAGGGCTTCCACGTGGACCACACGGCCGAGCTGCGGGGAATCCGGTGGACCAGCAGT TTGCGGCGGGAGACCTCGGACTATCACCGCACGCTGACGCCCCACCCTGGAGGCACTGTTTGTAGTAGTT TTCAGAAGACAGAGTTAGAGGCAAGCTGCGTGGGTGCTCGGTACTGAATTATAGGATGGGAACCTCCAG TGTCTCTGATACATTTCCAGCTGCACCTTCTGCTGCGACCCCTCCAGACGCTGAGCCTGGGCCTGGAGGAG GAGCTATTGCAGGAGGGATCCGGGCAAGGCTGCGGGAGCACGGCATCTCCCTGGCTGCCTATGCCACAA </p>

TTGTGTCGGCTGAGCTCACAGGGAGACATAAGGGACCCTTGGCAGAAAGAGACTTCAAATCAGGCCGCTG
TCCAGGGAACTCCTTTTCTGCGGGAACAGCCAGTGTGTGACCAAGGTGAACCCGGAGTGTGACGACCAG
GAGGACTGCTCCGATGGGTCCGACGAGGCGCACTGCGAGTGTGGCTTGCAGCCTGCCTGGAGGATGGCCG
GCAGGATCGTGGGCGGCATGGAAGCATCCCCGGGGGAGTTTCCGTGGCAAGCCAGCCTTCGAGAGAACAA
GGAGCACTTCTGTGGGGCCGCATCATCAACGCCAGGTGGCTGGTGTCTGCTGCTCACTGCTTCAATGAG
TTCCAAGACCCGACGAAGTGGGTGGCTTACGTGGGTGCGACCTACCTCAGCGGCTCGGAGGCCAGCACCG
TGCGGGCCCAGGTGGTCCAGATCGTCAAGCACCCCTGTACAACGCGGACACGGCCGACTTTGACGTGGC
TGTGCTGGAGCTGACCAGCCTCTGCCTTTCCGGCCGACATCCAGCCCGTGTGCCTCCCGGCTGCCACA
CACATCTTCCACCCAGCAAGAAGTGCCTGATCTCAGGCTGGGGCTACGTGCTGCAGAAAGCCACTGTGG
AGCTGCTGGACAGGCACGTGTGCCAGCTTGTACGGCCATCACTCACTGACAGGATGGTGTGCGCTGG
CTACCTGGACGGGAAGGTGGACTCCTGCCAGGGTGAAGTCAAGGAGGCCCTGGTCTGCCAGGAGCCCTCT
GGCCGGTTGTTTCTGGCTGGCATCGTGAGCTGGGGAATCGGGTGTGCGGAAGCCCGGCATCCAGGGGTCT
ATGCCCGAGTCACAGGCTACGCGACTGGATCCTGGAGGCCACCACAAAGCCAGCATGCCTCTGGCCCC
CACCATGGCTCCTGCCCTGCCGCCCCCAGCACAGCCTGGGCCACCAGTCTGAGAGCCCTGTGGTTCAGC
ACCCCAACCAATCGATGCAGGCCCTCAGTACCGTGCCTCTTGAAGTGGTCAACCGTTCCTAAGCTACAAG
GTATTTTCGGGGCAGAAAGGTAGAAAGATGATGTACGTGCCTATCTTGATTTAGGGAGAACGGATATCGTC
ATAGTATCTTCATAATTTTGGATCTTCTGTTCAAGGAAAGGTACATGTGTATCCGTTTATTCCTATCT
TACGTGCGTGTACCTCATGGTATCT

The disclosed NOV23b nucleic acid sequence, located on chromosome 19, has 208 of 327 bases (63%) identical to a gb:GENBANK-ID:AF098327|acc:AF098327.1 mRNA from *Homo sapiens* (putative mast cell mMCP-7-like II typtase gene, complete cds) ($E = 2.8e^{-14}$).

A disclosed NOV23b polypeptide (SEQ ID NO:100) encoded by SEQ ID NO:99 is 510 amino acid residues and is presented using the one-letter amino acid code in Table 23D. Signal P, Psort and/or Hydropathy results predict that NOV23b contains no signal peptide and is likely to be localized in the microbody (peroxisome) with a certainty of 0.5131.

Table 23D. Encoded NOV23b protein sequence (SEQ ID NO:100).

MGRHEQDLFLPSRKTRGHLGLTWHSSTLCCTWGPPAFLSTQGFHVDTAELRGIRWTSLSRRETS DYHRTLTP
LEALFVSSFQKTELEASCVGC SVLN YRDGNSSVLVHFQ LHFLLRPLQTL SLGLEELLQ R G I R A R L R E H G I S L A A
YGTIVSAELTGRHKGPLAERDFKSGRC PGNSF SCGNSQCVTKVNPECDDQEDCS DGSDEAHCECGLQPAWRMAGR
IVGMEASPGEF PWQASLRENKEHFCGAAI INARWLVSAAHCFNEFQDPTK W V A Y V G A T Y L S G S E A S T V R A Q V V Q
IVKHPLYNADTADFDVAVLELTSPLFPGRHIQPVCLPAATHIFPPSKKCLISGWGYVLQKATVELLDQALCASLY
GHSLTDRMVCAGYLDGKVDSCQDGGGGLVCEPSPGRFLAGIVSWGIGCAEARHPGVYARVTRLRDWILEATTK
ASMLAPTMAPAPAAPSTAWPTSPESPVVSTPTKSMQALSTVPLDWVTPKLQGFIGAER

The disclosed NOV23b amino acid sequence has 109 of 246 amino acid residues (44%) identical to, and 152 of 246 amino acid residues (61%) similar to, the 422 amino acid residue ptnr:SPTREMBL-ACC:Q9UL52 protein from *Homo sapiens* (Human) (serine protease DESC1) ($E = 1.3e^{-55}$).

NOV23b is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain-amygdala, brain-cerebellum, brain-hippocampus, brain-substantia nigra, brain-thalamus, brain-whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma-Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the NOV23b sequence.

NOV23c

A disclosed NOV23c nucleic acid of 815 nucleotides (also referred to as CG56647-01) encoding a novel adrenal secretory serine protease-like protein is shown in Table 23E. An open reading frame was identified beginning with a GGT initiation codon at nucleotides 1-3 and ending with a TAA codon at nucleotides 787-789. The start and stop codons are shown in bold in Table 23E, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon of NOV23c is not a traditional initiation codon, NOV23c could be a partial reading frame that extends further in the 5' direction.

Table 23E. NOV23c nucleotide sequence (SEQ ID NO:101).

GGTCCTGCCTTCGTGGGGTCATGGCTGGTGACCTGCTGTCTTGCAGAGTGTGGCTTGCAGCCTGCCTGGAGG
ATGGCCGGCAGGATCGTGGGCGGCATGGAAGCATCCCCGGGGAGTTCCGTGGCAAGCCAGCCTTCGAGAG
AACAAGGAGCACTTCTGTGGGGCCGCCATCAACGCCAGGTGGCTGGTGTCTGCTGCTCACTGCTTCAAT
GAGTTCCAAGACCCGACGAAGTGGGTGGCCTACGTGGGTGCCACCTACCTCAGCGGCTCGGAGGCCAGCACC
GTGCGGGCCAGGTGGTCCAGATCGTCAAGCACCCCTGTACAACGCGGACACGGCCGACTTTGACGTGGCT
GTGCTGGAGCTGACCAGCCCTCTGCCTTTTCGGCCGGCAGATCCAGCCCGTGTGCCTCCCGCTGCCACACAC
ATCTTCCCACCCAGCAAGAAGTGCCTGATCTCAGGCTGGGGCTACCTCAAGGAGGACTTCGTAAGCATCTT
CCTCTGCAGAAAGCCACTGTGGAGCTGCTGGACCAGGCACTGTGTGCCAGCTTGTACGGCCATTCACTCACT
GACAGGATGGTGTGCGCTGGCTACCTGGACGGGAAGGTGGACTCCTGCCAGGGTGACTCAGGAGGACCCCTG
GTCTGCGAGGAGCCCTCTGGCCGGTCTTTCTGGCTGGCATCGTGAGCTGGGGAATCGGGTGTGCGGAAGCC
CGCGTCCAGGGGTCTATGCCCGAGTCACAGGCTACGTGACTGGATCCTGGAGGCCACCCGTTCCT**AAGCT**
ACAAGGTATTTTCGGGGCAGAAA

In a search of public sequence databases, the NOV23c nucleic acid sequence, located on chromosome 19, has 350 of 564 bases (62%) identical to a gb:GENBANK-ID:E13204|acc:E13204.1 mRNA from *Homo sapiens* (Human cDNA encoding a serine protease) (E = 3.2e⁻²⁶).

The disclosed NOV23c polypeptide (SEQ ID NO:102) encoded by SEQ ID NO:101 has 262 amino acid residues and is presented in Table 23F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV23c has no signal peptide and is likely to be localized extracellularly with a certainty of 0.3750. Alternatively, NOV23c may also localize to the microbody (peroxisome) with a certainty of 0.1391, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV23c is between positions 15 and 16: CLA-EC.

Table 23F. Encoded NOV23c protein sequence (SEQ ID NO:102).

GPAFVGSWLVTCCLAECGLQPAWRMAGRIVGGMEASPGFPPWQASLRENKEHFCGAAIINARWLVSAAHCFN
EFQDP TKWVAYVGATYLSGSEASTVRAQVVQIVKHPLYNADTADFVAVLELTSPLPFRHIQPVCLPAATH
IFPPSKKCLISGWYGLKEDFRKHLPLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVDSCQGDGGPL
VCEEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWILEATRS

A search of sequence databases reveals that the NOV23c amino acid sequence has 114 of 248 amino acid residues (45%) identical to, and 152 of 248 amino acid residues (61%) similar to, the 273 amino acid residue ptnr:TREMBLNEW-ACC:BAB20278 protein from *Mus musculus* (Mouse) (Type 1 Spinesin) ($E = 1.1e^{-53}$).

NOV23c is predicted to be expressed in at least the following tissues: Ovary, kidney, breast, lung, muscle, liver, spleen, blood, lymphocyte. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

The disclosed NOV23a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 23G.

Table 23G. BLAST results for NOV23					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12836503 dbj BAB23684.1 (AK004939)	data source:SPTR, source key:O95519, evidence:ISS~homolog to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT)~putative [<i>Mus musculus</i>]	799	136/280 (48%)	180/280 (63%)	1e-75
gi 16758444 ref NP_446087.1 (NM_053635)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [<i>Rattus norvegicus</i>]	855	133/289 (46%)	182/289 (62%)	3e-72
gi 10257390 gb AAG15395.1 AF057145_1 (AF057145)	serine protease TADG15 [<i>Homo sapiens</i>]	855	132/289 (45%)	185/289 (63%)	3e-72
gi 11415040 ref NP_068813.1 (NM_021978)	suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [<i>Homo sapiens</i>]	855	132/289 (45%)	185/289 (63%)	3e-72
gi 12249015 dbj BAB20376.1 (AB030036)	prostamin [<i>Homo sapiens</i>]	855	131/289 (45%)	184/289 (63%)	9e-72

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 23H. In the ClustalW alignment of the NOV23 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate

regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 23H. ClustalW Analysis of NOV23

- 1) Novel NOV23a (SEQ ID NO:98)
- 2) Novel NOV23b (SEQ ID NO:100)
- 2) Novel NOV23c (SEQ ID NO:102)
- 4) gi|12836503|dbj|BAB23684.1| (AK004939) data source:SPTR, source key:095519, evidence:ISS-homolog to DJ1170K4.4 (NOVEL PROTEIN) (FRAGMENT)-putative [*Mus musculus*] (SEQ ID NO:416)
- 5) gi|16758444|ref|NP_446087.1| (NM_053635) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin) [*Rattus norvegicus*] (SEQ ID NO:417)
- 6) gi|10257390|gb|AAG15395.1|AF057145_1 (AF057145) serine protease TADG15 [*Homo sapiens*] (SEQ ID NO:418)
- 7) gi|11415040|ref|NP_068813.1| (NM_021978) suppression of tumorigenicity 14 (colon carcinoma, matriptase, epithin); suppression of tumorigenicity 14 (colon carcinoma); matriptase [*Homo sapiens*] (SEQ ID NO:419)
- 8) gi|12249015|dbj|BAB20376.1| (AB030036) prostamin [*Homo sapiens*] (SEQ ID NO:420)

			10	20	30	40	50	60	
25	NOV27a	1						
	NOV27b	1	-----MCDRHEQD-----LFLPS-----RKTRGHLGLT-						23
	NOV27c	1	-----MCDRHEQD-----LFLPS-----RKTRGHLGLT-						23
	gi 12836503	1	MPTTEVPQAADGGQDAGDGEAAE-----PEGKFKPPK-----NTKRKNRDYVRF						47
	gi 16758444	1	MGNNRGRKAGGGSQDFGAGLKYNSRLENMNGFEEGVEFLPVNNAKQVEKRGPRRWVVA						60
	gi 10257390	1	MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVLA						60
30	gi 11415040	1	MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVLA						60
	gi 12249015	1	MGSDRARKGGGGPKDFGAGLKYNSRHEKVNGLEEGVEFLPVNNVKKVEKHGPGRWVLA						60
			70	80	90	100	110	120	
35	NOV27a	23						
	NOV27b	23	-----WHSS-----TLCCTWG-----PPAFLS-----						40
	NOV27c	1	-----WHSS-----TLCCTWG-----PPAFLS-----						40
	gi 12836503	48	LLL-VLAALVSAGVMLWYFLGYKAEVTSQVYSGSLRVLNRHFSQDLGRRESIAFRSESA						106
	gi 16758444	61	VVFSFLLLSLMAGLLVWHFHYR--NVRIQKVFNGHLRITNENFLDAYENSSTEFVSLAS						118
40	gi 10257390	61	VLIGLLLVLGIGFLVWHLQYR--DVRVQKVFNGYMRITNENFVDAYENSSTEFVSLAS						118
	gi 11415040	61	VLIGLLLVLGIGFLVWHLQYR--DVRVQKVFNGYMRITNENFVDAYENSSTEFVSLAS						118
	gi 12249015	61	VLIGLLLVLGIGFLVWHLQYR--DVRVQKVFNGYMRITNENFVDAYENSSTEFVSLAS						118
			130	140	150	160	170	180	
45	NOV27a	40						
	NOV27b	40	-----TOGFHVDHTAELR-----GIRWTSS-----LRRETSDYHR-TLTP-TLEALF						80
	NOV27c	1	-----TOGFHVDHTAELR-----GIRWTSS-----LRRETSDYHR-TLTP-TLEALF						80
	gi 12836503	107	KAQKMLQELVASTRL-GTYYNSSSVYSFGEGLPTCFWFILDIPYQRLTSPPEVVR						165
50	gi 16758444	119	QVKEALKLMYSEPVLCGPYHKKSTVTAFSEGSVIAYYWSEFSIPPHLEEEVDRA						178
	gi 10257390	119	KVKDALKLLYSGVPFLGPYHKBSAVTAFSEGSVIAYYWSEFSIPOHLVEEAERVMA						178
	gi 11415040	119	KVKDALKLLYSGVPFLGPYHKBSAVTAFSEGSVIAYYWSEFSIPOHLVEEAERVMA						178
	gi 12249015	119	KVKDALKLLYSGVPFLGPYHKBSAVTAFSEGSVIAYYWSEFSIPOHLVEEAERVMA						178
55			190	200	210	220	230	240	
	NOV27a	81						
	NOV27b	81	-----VSSFQ-----KTE-----LEASCVGC SVLN-----YR-----						102
	NOV27c	1	-----VSSFQ-----KTE-----LEASCVGC SVLN-----YR-----						102
60	gi 12836503	166	VDELLSNSSTLASVKTEYEVDPEGLVLEASVNDIVVLNSTLGCYRYSYVNPQVLP						225
	gi 16758444	179	VTLPPTARALKSFVLTISVVAFFIDPRMLQRTQDNSCSFALHARGRTVTRFTTPG--						236

	gi 10257390	179	VMLPPRARSLKSFVVTSVVAFPDTSKTVQRTQDNSSCSFGLHARGVELMRFTTPG--FPDS	236
	gi 11415040	179	VMLPPRARSLKSFVVTSVVAFPDTSKTVQRTQDNSSCSFGLHARGVELMRFTTPG--FPDS	236
	gi 12249015	179	VMLPPRARSLKSFVVTSVVAFPDTSKTVQRTQDNSSCSFGLHARGVELMRFTTPG--FPDS	236
5			250 260 270 280 290 300	
	NOV27a	102	---DGNSSVLVHFQ-----	113
	NOV27b	102	---DGNSSVLVHFQ-----	113
	NOV27c	1	-----	1
10	gi 12836503	226	PDQQTTSCLWHLQGPEDLMIKVRLWTRVDCRDR----VAMYDAAGPLEKRLITSVYGC	280
	gi 16758444	237	PYPAHARCQWVLRGDADSVLSLTFRSFDVAPCDGHSDLVTVYDSLSPMEPHAVVRLCGT	296
	gi 10257390	237	PYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLTVYNTLSPMEPHALVQLCGT	296
	gi 11415040	237	PYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLTVYNTLSPMEPHALVQLCGT	296
	gi 12249015	237	PYPAHARCQWALRGDADSVLSLTFRSFDLASCDERGSDLTVYNTLSPMEPHALVQLCGT	296
15			310 320 330 340 350 360	
	NOV27a	113	-----LH-----FLLRP-----LQ	122
	NOV27b	113	-----LH-----FLLRP-----LQ	122
	NOV27c	1	-----	1
20	gi 12836503	281	SRQEPVMEVLASGSVMVWKKGMHSYYDPFLLSVKSVAFQDCQVNLTLLEGLDQTQGFRL	340
	gi 16758444	297	FSPSYNTLFLSSQNVLVTLTINTDRRHPCGEATFFQLPKMSSCGG---LLSEAQGTFS	352
	gi 10257390	297	YPPSYNTLFHSSQNVLITLTINTERRHPCGEATFFQLPRMSSCGG---RLRKAQGTFN	352
	gi 11415040	297	YPPSYNTLFHSSQNVLITLTINTERRHPCGEATFFQLPRMSSCGG---RLRKAQGTFN	352
	gi 12249015	297	YPPSYNTLFHSSQNVLITLTINTERRHPCGEATFFQLPRMSSCGG---RLRKAQGTFN	352
25			370 380 390 400 410 420	
	NOV27a	123	-----LSLGLLEEELLRGIRARLREHG-----	145
	NOV27b	123	-----LSLGLLEEELLRGIRARLREHG-----	145
	NOV27c	1	-----	1
	gi 12836503	341	TPPYPSYSPSTHCSWHLTVPSLDYGLALWFDAYALRRQKYNRLCTQGQWMIQNRRLCGF	400
	gi 16758444	353	SPYYPGHYPPNINCTWNIKVPNNRNKVRKLFYLVDPNIPVGSCTKDYEINGEKFCGE	412
	gi 10257390	353	SPYYPGHYPPNIDCTWNIQVNNQHVKVSFKFFYLLEPCVPAGTCKPDYEINGEKYCGE	412
	gi 11415040	353	SPYYPGHYPPNIDCTWNIQVNNQHVKVSFKFFYLLEPCVPAGTCKPDYEINGEKYCGE	412
	gi 12249015	353	SPYYPGHYPPNIDCTWNIQVNNQHVKVSFKFFYLLEPCVPAGTCKPDYEINGEKYCGE	412
30			430 440 450 460 470 480	
	NOV27a	145	-----ISLAAYGTIVSAELT-----	160
	NOV27b	145	-----ISLAAYGTIVSAELT-----	160
	NOV27c	1	-----	1
	gi 12836503	401	RTLQPYAERIPMVASDGVITNFTSQISLTGPGVQVYYSLYNQSDPCPGFELCSV-----	454
	gi 16758444	413	RS-----QFVVSSNSSKITVRFHSDQSYTDTGFLAEYLSYDSNDPCPGMFMCKTGRCIRK	467
	gi 10257390	413	RS-----QFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTCRTGRCIRK	467
	gi 11415040	413	RS-----QFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTCRTGRCIRK	467
	gi 12249015	413	RS-----QFVVTSNSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRTCRTGRCIRK	467
35			490 500 510 520 530 540	
	NOV27a	160	-----GRHKGP--LAER-----	170
	NOV27b	160	-----GRHKGP--LAER-----	170
	NOV27c	1	-----	1
	gi 12836503	454	-----NGLCVPACDGIKDCPNGLDERNVCVRA	481
	gi 16758444	468	DLRCDGWADCPDYSERHRCRNATHQFMCKNQFCKPLFWVCDVNDGCGDSDEEGCSCPA	527
	gi 10257390	468	ELRCDGWADCTDHSDELNCSDAGHQFTCKNKFKPLFWVCDVNDGCDNSDEQGCSCPA	527
	gi 11415040	468	ELRCDGWADCTDHSDELNCSDAGHQFTCKNKFKPLFWVCDVNDGCDNSDEQGCSCPA	527
	gi 12249015	468	ELRCDGWADCTDHSDELNCSDAGHQFTCKNKFKPLFWVCDVNDGCDNSDEQGCSCPA	527
40			550 560 570 580 590 600	
	NOV27a	170	-----DFKSGR-----CPGNSFSCGNSQCQVTKVNPEC	197
	NOV27b	170	-----DFKSGR-----CPGNSFSCGNSQCQVTKVNPEC	197
	NOV27c	1	-----GPAFVG--SWLVT-----C	12
	gi 12836503	482	MFQCQEDSTCISLPRVCDRQPDCLNGSDEEQCQEG--VPCGTFTFOCEDRSCVKKKNPEC	539
	gi 16758444	528	GSFKCSNGKCLPQSQQCNGKDDCGDGSDEASCDNVNAVSTKYTYRCNGGLCNKGNPEC	587
	gi 10257390	528	QTFRCSTNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVTCTKHTRYRCLNGLCLSKGNPEC	587
	gi 11415040	528	QTFRCSTNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVTCTKHTRYRCLNGLCLSKGNPEC	587
	gi 12249015	528	QTFRCSTNGKCLSKSQQCNGKDDCGDGSDEASCPKVNVTCTKHTRYRCLNGLCLSKGNPEC	587
60				
65				
70				

		610	620	630	640	650	660								
	NOV27a	198	DDQEDCS	DGSD	EAHCE	CGLOPA	WRMAGRI	VGGMEAS	PG	GEFPWQ	ASLREN	KE-HF	CGAA	II	256
	NOV27b	198	DDQEDCS	DGSD	EAHCE	CGLOPA	WRMAGRI	VGGMEAS	PG	GEFPWQ	ASLREN	KE-HF	CGAA	II	256
5	NOV27c	12	-----	-----	CLAE	CGLOPA	WRMAGRI	VGGMEAS	PG	GEFPWQ	ASLREN	KE-HF	CGAA	II	59
	gi 12836503	540	DGQSDCR	DGSD	EQHDC	CGLO	---	GLSSRI	VGGT	VSSEGE	WPWQAS	LQIRGR	-HIC	CGG	595
	gi 16758444	588	DGKKDCS	DGSD	EKNDC	D	CGLR-S	FTKQAR	VVG	TNADEGE	WPWQV	SLHALG	QGHLC	GASLI	646
	gi 10257390	588	DGKEDCS	DGSD	EKNDC	D	CGLR-S	FTKQAR	VVG	TDADEGE	WPWQV	SLHALG	QGHLC	GASLI	646
10	gi 11415040	588	DGKEDCS	DGSD	EKNDC	D	CGLR-S	FTKQAR	VVG	TDADEGE	WPWQV	SLHALG	QGHLC	GASLI	646
	gi 12249015	588	DGKEDCS	DGSD	EKNDC	D	CGLR-S	FTKQAR	VVG	TDADEGE	WPWQV	SLHALG	QGHLC	GASLI	646
		670	680	690	700	710	720								
15	NOV27a	257	NARWLVS	AAHCFNE	-----	FODPTK	WVAVV	GATYLS	GSEAST	VR-AQ	VVQIVK	HPLYN	AD	310	
	NOV27b	257	NARWLVS	AAHCFNE	-----	FODPTK	WVAVV	GATYLS	GSEAST	VR-AQ	VVQIVK	HPLYN	AD	310	
	NOV27c	60	NARWLVS	AAHCFNE	-----	FODPTK	WVAVV	GATYLS	GSEAST	VR-AQ	VVQIVK	HPLYN	AD	113	
	gi 12836503	596	ADRWVIT	AAHCFQ	EDS---	MASPKL	WTVF	ECKMR	NSRWPG	EVSKF	VSRFLF	HPHYE	ED	651	
	gi 16758444	647	SPDWLVS	AAHCFQ	DETIF	KYS	DEHMT	WTAFL	GLLDQ	SKRSAS	GVQEHK	LKRIT	TH	706	
	gi 10257390	647	SPNWLVS	AAHCFQ	DETIF	KYS	DEHMT	WTAFL	GLLDQ	SKRSAS	GVQEHK	LKRIT	TH	706	
20	gi 11415040	647	SPNWLVS	AAHCFQ	DETIF	KYS	DEHMT	WTAFL	GLLDQ	SKRSAS	GVQEHK	LKRIT	TH	706	
	gi 12249015	647	SPNWLVS	AAHCFQ	DETIF	KYS	DEHMT	WTAFL	GLLDQ	SKRSAS	GVQEHK	LKRIT	TH	706	
		730	740	750	760	770	780								
25	NOV27a	311	TADFVAV	LELTS	PLP	EGRHI	QPVCL	PAATHI	FPPSK	KCLIS	GWG	YLKED	FV	370	
	NOV27b	311	TADFVAV	LELTS	PLP	EGRHI	QPVCL	PAATHI	FPPSK	KCLIS	GWG	Y-----	VLQ	360	
	NOV27c	114	TADFVAV	LELTS	PLP	EGRHI	QPVCL	PAATHI	FPPSK	KCLIS	GWG	YLKED	FR	172	
	gi 12836503	652	SHDYDVA	LLQLD	HEFV	VYSAT	VRP	CLPAR	SHFF	EPGQ	HCWIT	GWGAQ	REGG	710	
	gi 16758444	707	TFDYDIA	LLLELEK	PAEY	STV	VRP	ICLPD	NTHV	FPAG	KAIW	VTGW	GHTKEG	765	
30	gi 10257390	707	TFDYDIA	LLLELEK	PAEY	SSM	VRP	ICLPD	ASHV	FPAG	KAIW	VTGW	GHTQYGG	765	
	gi 11415040	707	TFDYDIA	LLLELEK	PAEY	SSM	VRP	ICLPD	ASHV	FPAG	KAIW	VTGW	GHTQYGG	765	
	gi 12249015	707	TFDYDIA	LLLELEK	PAEY	SSM	VRP	ICLPD	ASHV	FPAG	KAIW	VTGW	GHTQYGG	765	
		790	800	810	820	830	840								
35	NOV27a	371	ATVELLD	QALCAS	LYGH	SLTDR	MVCAG	YLDG	KVDSC	QGD	SGG	PLV	CEE	430	
	NOV27b	361	ATVELLD	QALCAS	LYGH	SLTDR	MVCAG	YLDG	KVDSC	QGD	SGG	PLV	CEE	420	
	NOV27c	173	ATVELLD	QALCAS	LYGH	SLTDR	MVCAG	YLDG	KVDSC	QGD	SGG	PLV	CEE	232	
40	gi 12836503	711	VDVQLV	PDLC	SEAY	RYQ	VS	PRML	CAGY	RK	GK	KDAC	QGD	770	
	gi 16758444	766	GEIRVIN	OTTCE	ELL	PQOI	TPRM	MCV	GEL	SGG	VDS	CQGD	SGG	825	
	gi 10257390	766	GEIRVIN	OTTCE	NLL	PQOI	TPRM	MCV	GEL	SGG	VDS	CQGD	SGG	825	
	gi 11415040	766	GEIRVIN	OTTCE	NLL	PQOI	TPRM	MCV	GEL	SGG	VDS	CQGD	SGG	825	
	gi 12249015	766	GEIRVIN	OTTCE	NLL	PQOI	TPRM	MCV	GEL	SGG	VDS	CQGD	SGG	825	
		850	860	870	880	890	900								
45	NOV27a	431	WGIGCA	EARR	PGVY	ARV	TR	LRD	WILE	A-----	-----	-----	-----	460	
	NOV27b	421	WGIGCA	EARR	PGVY	ARV	TR	LRD	WILE	A	TKAS	MPLA	PTMA	480	
	NOV27c	233	WGIGCA	EARR	PGVY	ARV	TR	LRD	WILE	A	-----	-----	-----	262	
50	gi 12836503	771	WGLG	GRPN	FFGV	YTR	VTR	VIN	WIO	QVLT	-----	-----	-----	799	
	gi 16758444	826	WEGGCA	QRN	KPG	VYTR	IPE	VRD	WIK	EOT	GV-----	-----	-----	855	
	gi 10257390	826	WGDGCA	QRN	KPG	VYTR	IPL	FRD	WIK	E	NTGV-----	-----	-----	855	
	gi 11415040	826	WGDGCA	QRN	KPG	VYTR	IPL	FRD	WIK	E	NTGV-----	-----	-----	855	
55	gi 12249015	826	WGDGCA	QRN	KPG	VYTR	IPL	FRD	WIK	E	NTGV-----	-----	-----	855	
		910	920	930											
	NOV27a	460	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	460	
60	NOV27b	481	TPTKSM	QALST	VP	LD	WV	TV	PK	LQ	GIF	GA	ER	510	
	NOV27c	262	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	262	
	gi 12836503	799	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	799	
	gi 16758444	855	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	855	
	gi 10257390	855	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	855	
	gi 11415040	855	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	855	
65	gi 12249015	855	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	855	

Tables 23I-L list the domain descriptions from DOMAIN analysis results against NOV23. This indicates that the NOV23 sequence has properties similar to those of other proteins known to contain this domain.

Table 23I Domain Analysis of NOV23a

gnl|Smart|smart00020, Tryp_SPC, Trypsin-like serine protease; Many of these are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. (SEQ ID NO:812)
CD-Length = 230 residues, 100.0% aligned
Score = 269 bits (687), Expect = 3e-73

5	NOV23:	225	RIVGGMEASPGFEPWQASLR-ENKEHFCGAAIINARWLVSAAHCFNEFQDPTKWVAYVGA	283
	Sbjct:	1	RIVGGSEANIGSFPPWQVSLQYRGGRHFCGGSLSIPRWLTAACHCVYGSA-PSSIRVRLGS	59
10	NOV23:	284	TYLSGSEASTVRAQVVQIVKHPLYNADTADFDVAVLELTSPLPFGRHIQPVCLPAATHIF	343
	Sbjct:	60	HDLSSGEETQTV-KVSKVIVHPNYPSTYDNDIALLLKLSEPVTLSDTVRPICLPSSGYNV	118
15	NOV23:	344	PPSKKCLISGWGYLKEDFVVKPEVLQKATVELLDQALCASLY--GHSLTDRMVCAGYLDG	401
	Sbjct:	119	PAGTTCVSGWGRTSESSGSLPDTLQEVNVPVSNATCRRAYSGGPAITDNMLCAGGLEG	178
20	NOV23:	402	KVDSCQGDSSGGLVCEEPSGRFFLAGIVSWG-IGCAEARRPGVYARVTRLRDWI	454
	Sbjct:	179	GKDACQGDSSGGLVCN--DPRWVLVGIVSWGSGYCARPNKPGVYTRVSSYLDWI	230

Table 23J Domain Analysis of NOV23a

gnl|Pfam|pfam00089, trypsin, Trypsin. Proteins recognized include all proteins in families S1, S2A, S2B, S2C, and S5 in the classification of peptidases. Also included are proteins that are clearly members, but that lack peptidase activity, such as haptoglobin and protein Z (PRTZ*). (SEQ ID NO:813)
CD-Length = 217 residues, 100.0% aligned
Score = 223 bits (568), Expect = 2e-59

25	NOV27:	226	IVGMEASPGFEPWQASLRENKEHFCGAAIINARWLVSAAHCFNEFQDPTKWVAYVGATY	285
	Sbjct:	1	IVGGREAQAGSFPPWQVSLQVSSGHFCGGSLSIENWVLTAACHCVS---GASSVRVVLGEHN	57
30	NOV27:	286	LSGSEASTVRAQVVQIVKHPLYNADTADFDVAVLELTSPLPFGRHIQPVCLPAATHIFPP	345
	Sbjct:	58	LGTTEGTEQKFDVKKIIVHPNYPDT--NDIALLLKLSPVTLGDTVRLPCLPSASSDLFV	115
35	NOV27:	346	SKKCLISGWGYLKEDFVVKPEVLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVDS	405
	Sbjct:	116	GTCVSVSGWGRTKNL--GTSDTLQEVVVPVSVRETCSRAYSAGGTVDTMICAGALGGK-DA	172
	NOV27:	406	CQGDSSGGLVCEEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWI	454
	Sbjct:	173	CQGDSSGGLVCSGDG---ELVGIVSWGYGCAVGNYPGVYTRVSRYLDWI	217

THE UNIVERSITY OF CHICAGO PRESS

5

10

gnl|Pfam|pfam00057, ldl_recept_a, Low-density lipoprotein receptor
domain class A (SEQ ID NO:815)
CD-Length = 39 residues, 94.9% aligned
Score = 43.5 bits (101), Expect = 3e-05

15

20

25

sequence of 3,089 daltons and a 245-amino acid catalytic region of 27,458 daltons. The amino acid sequence of beta-tryptase was found to be 90% identical with that of alpha-tryptase, the first 20 amino acids of the catalytic portions being 100% identical. Both alpha- and beta-tryptase sequences were localized to human chromosome 16 by analysis of DNA preparations from 25 human/hamster somatic cell hybrids by PCR.

Because of the presence of the trypsin domains and the homology to the adrenal secretory serine protease, it is anticipated that the novel sequences described here will have useful properties and functions similar to these proteins.

The disclosed NOV23 nucleic acid of the invention encoding an Adrenal secretory serine protease -like protein includes the nucleic acids whose sequences are provided in Tables 23A, 23C, 23E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 23A, 23c, or 23E while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV23 protein of the invention includes the Adrenal secretory serine protease -like protein whose sequence are provided in Table 23B, 23D, or 23F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 23B, 23D, or 23F while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 55 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that these Adrenal secretory serine protease - like proteins (NOV23) is a member of a "Adrenal secretory serine protease family".

Therefore, the NOV23 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV23 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, endometriosis, fertility, anemia, ataxia-telangiectasia, autoimmune disease, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphoedema, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, and/or other diseases and pathologies.

NOV23 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV23 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV23 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV24

NOV24 includes two novel parchorin-like proteins disclosed below. The disclosed sequences have been named NOV24a and NOV24b.

NOV24a

A disclosed NOV24a nucleic acid of 2091 nucleotides (also referred to as CG56455-01) encoding a novel parchorin-like protein is shown in Table 24A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 7-9 and ending with a TGA codon at nucleotides 2080-2082. The start and stop codons are shown in bold in Table 24A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 24A. NOV24a nucleotide sequence (SEQ ID NO:103).

GCGGCCATGGCCGAGGCCGCGGAGCCGAGGGGGTTGCCCGGGTCCCCAGGGGCCGCGGAGGTCCCCGCG
 CCTCTGGCTGAGAGACCCGAGAGCCAGGAGCCGCGGGCGGGGAGGCAGAAGGGCGAGGGGAGCGAGGGC
 GCAGAGGAGGCGCCGAGGGGCGCCCGCTGTGAAGGAGGCAGGAGGCGGGCCAGACAGGGGCCCGGAG
 GCCGAGGCGCGGGGACGAGGGGGGCGCACGGCGAGACTGAGGCCGAGGAGGGAGCCCCGAGGGGTGCCGAG
 GTGCCCAAGGAGGGGAGGAGACAAGCGGCGCGCAGCAGGTGGAGGGGGCGAGCCCCGGGACGCGGCGCGCAG
 GCGAGCCCCGCGGGGAGGCTCAGAGGGAGCCCGAGGACTCTGCGGCCCGGAGAGGCAGGAGGAGCGGAG
 CAGAGGCCTGAGCTCCCGAAGGTAGCGCGTCCGGGGAGGCGGGGACAGCGTAGACGCGGAGGGCCCCGCTG
 GGGGACAACATAGAAGCGAGGGGCCCGGGCGGCACAGCGTAGAGGCGGAGGGCCGGGTGGGGACAGCGTA
 GACGCGGAAGGTCCCGGGGGGACAGCGTAGACGCGGAGGGCCCCGCTGGGGGACAACATACAAGCCGAGGGC
 CCGCGGGGGGACAGCGTAGACGCGGAGGGCCGGGTGGGGGACAGCGTAGACGCGGAAGGTCCGGCGGGGAC
 AGCGTAGACGCGGAGGGCCGGGTGGGGGACAGCGTAGAGGCGGGGACCCGCGGGGGACGGCGTAGAAGCG
 GGGGTCCCGCGGGGGACAGCGTAGAAGCCGAAGGCCCGCGGGGGACAGCATGGACGCCGAGGGTCCGCA
 GGAAGGGCGCGCCGGGTCTCGGGTGAGCCGACGCAATCGGGGACGGCAGCCTCTCGGCCCGAGGCA
 ATTGAGGTGCGAGCCGGGAGAGTGGGGGCGCAGCCCCGGTGGCTCGCCTGGGACGCGAGCGGAGGCGG
 GAGTCCCGGGGTAAAGGGTCCGAAGAAGCGGCCCGGGGACGCAAGGGCAGACGCTGGCGAGGACAGG
 GTAGGGATGGGCCACAGCAGGAGCCGGGGAGGACGAAGAGAGACGAGAGCGGAGCCCGAGGGGCCAAGG
 GAGGAGCAAGCAGCGGGGGCGAAGAGGAATCCCCGACAGCAGCCACATGGGGAGGCCCTCAGGGGGCGCC
 GCGGAGCCTGAGGCCAGCTCAGCAACCACCTGGCCGAGGAGGGCCCCGCGAGGGTAGCGCGAGGCCGCG
 CGCGTGAACGCGCCCGGGAGGACGAGAGGCGTCCGAGCCCCGGGCCCTGGGGCAGGAGCAGACATCACC
 CTCTTCGTCAAGGCTGGTTATGATGGTGAGAGTATCGGAAATTGCCCGTCTTCTCAGCGTCTCTTTATGATT
 CTCTGGCTGAAAGGCGTTATATTTAATGTGACCACAGTGGACCTGAAAAGGAAACCCGAGACCTGCAGAAC
 CTGGCTCCCGGAACAAACCTCTCTTCATGACTTTTGATGGTGAAGTCAAGACGGATGTGAATAAGATCGAG
 GAGTTCTTAGAGGAGAAATTAGCTCCCCGAGGTATCCCAAGCTGGGGACCCAAATCCCCGAATCTAATTCC
 GCAGGAAATGACGTGTTTGCCAAATCTCAGCGTTTATAAAAAACACGAAGAAGGATGCAATGAGTTTCAT
 GAAAAGAACCTGCTGAAGGCCCTGAGGAAGCTGGATAATTACTTAAATAGCCCTCTGCTGATGAAATGAT
 GCCTACAGCACCGAGGATGTCACTGTTTCTGGAAGGAAGTTCTGGATGGGGACGAGCTGACGCTGGCTGAC
 TGCAACCTCTTACCCAAGCTCCATATTATTAAGGTTCTTCATTTTCAGATTGTGGCCAAGAGTACAGAGAT
 TTTGAATTTCTTCTGAAATGACTGGCATCTGGAGACTTGAATAATGCTTATGCTAGAGATGAGTTTCA
 AATACGTGTCCAGCTGATCAAGAGATTGAACACGCATATTAGATGTTGCAAAAAGAAATGAAATGAAGCTGG
 GCT

In a search of public sequence databases, the NOV24a nucleic acid sequence, located on chromosome 21, has 1347 of 1897 bases (71%) identical to a gb:GENBANK-
 ID:AB035520|acc:AB035520.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus*
 mRNA for parchorin, complete cds) ($E = 2.4e^{-175}$).

A disclosed NOV24a polypeptide (SEQ ID NO:104) encoded by SEQ ID NO:103 has 691 amino acid residues and is presented in Table 24B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV24a has no signal peptide and is likely to be localized to the nucleus with a certainty of 0.3000. Alternatively, NOV24a may also localize to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 24B. Encoded NOV24a protein sequence (SEQ ID NO:104).

MAEAAEPEGVAPGPQGPPEVPAPLAERPGEPAAGGEAEGPGESEGAEEAPRGAAVKEAGGGPDRGPEAE
 ARGTRGAHGETEAEEGAPEGAEVPGGEETSGAQQVEGASPRGAQGEPRGEAQREPEDSAAPERQEEAEQR
 PEVPEGSASGEAGDSVDAEGPLGDNIEAEPAGDSVEAEGRVGDSVDAEGPAGDSVDAEGPLGDNIEAEGPA
 GDSVDAEGRVGDSDAEGPAGDSVDAEGRVGDSEAGDPAGDVEAGVPAGDSVEAEGPAGDSMDAEGPAGR
 ARRVSGEPPQSGDGLSPQAEAEVAAGESAGRSPEGLAWDAEEAEVPGVKGSEEAAPGDARADAGEDRVG
 DGPQQEPGEDEERRERSPEGPREEEAAGGEEESPDSPPHGEASRGAAEPEAQLSNHLAEEGPAEGSGEAARV
 NGRREDGEASEPRALGQEHDTLTVKAGYDGESIGNCPFSQRLFMILWLKGVIFNVTTVDLKRKPADLQNL
 PGTNPPFTFDGEVKTDVNKEEFLEEKLAAPPYPKLGTQHPESNSAGNDVFAKFSAFIKNTKKDANEVHEK
 NLLKALRLDNYLNSPLPDEIDAYSTEDVTVSGRKFLEDGLDELTLADCNLLPKLHIIKVLHFQIVAKKYRDFE

FPSEMTGIWRYLNNAYARDEFNTCPADQEIEHAYSDVAKRMK

A search of sequence databases reveals that the NOV24a amino acid sequence has 414 of 655 amino acid residues (63%) identical to, and 453 of 655 amino acid residues (69%) similar to, the 637 amino acid residue ptrn:SPTREMBL-ACC:Q9N2G5 protein from

5 *Oryctolagus cuniculus* (Rabbit) (Parchorin) ($E = 2.5e^{-182}$).

NOV24a is predicted to be expressed in at least the following tissues: brain, lung, and kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

10 In addition, the sequence is predicted to be expressed in gastric parietal cells, choroid plexus, salivary duct, lacrimal gland, kidney, airway epithelia and chorioretinal epithelia because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AB035520|acc:AB035520.1) a closely related *Oryctolagus cuniculus* mRNA for parchorin, complete cds homolog.

15 **NOV24b**

A disclosed NOV24b nucleic acid of 859 nucleotides (also referred to as CG56455-02) encoding a novel parchorin-like protein is shown in Table 24C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 853-855. The start and stop codons are shown in bold in Table 24A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 24C. NOV24b nucleotide sequence (SEQ ID NO:105).

ATGGCCGAGGCCGCGGAGCCTGAGGCCAGCTCAGCAACCACCTGGCCGAGGAGGGCCCCGCCGAGGGTAGC
GGCGAGGCCGCGCGTGTGAACGGCCGCGGGAGGACGGAGAGGCGTCCGAGCCCCGGGCCCTGGGGCAGGAG
CACGACATCACCCCTCTTCGTCAAGGCTGGTTATGATGGTGAGAGTATCGGAAATGCCCCGTTTCTCAGCGT
CTCTTTATGATTCTCTGGCTGAAAGGCGTTATATTTAATGTGACCACAGTGGACCTGAAAAGGAAACCCGCA
GACCTGCAGAACCTGGCTCCCGGAACAAACCCCTCCTTTTATGACTTTTGATGGTGAAGTCAAGACGGATGTG
AATAAGATCGAGGAGTTCTTAGAGGAGAAATTAGCTCCCCGAGGTATCCCAAGCTGGGGACCCAACATCCC
GAATCTAATTCCGACGAAATGACGTGTTTGCCAAATTCTCAGCGTTTATAAAAAACACGAAGAAGGATGCA
AATGAGATTGATGAAAAGAACCTGCTGAAGGCCCTGAGGAAGCTGGATAATTACTTAAATAGCCCTCTGCCT
GATGAAATAGATGCCTACAGCACCGAGGATGTCACTGTTTCTGGAAGGAAGTTTCTGGATGGGGACGAGCTG
ACGCTGGCTGACTGCAACCTCTTACCCAAGCTCCATATTATTAAGATTGTGGCCAAGAAGTACAGAGATTTT
GAATTTCTTCTGAAATGACTGGCATCTGGAGATACTTGAATAATGCTTATGCTAGAGATGAGTTCACAAAT
ACGTGTCCAGCTGATCAAGAGATTGAACACGCATATTCAGATGTTGCAAAAAGAAATGAAATGAAGCT

In a search of public sequence databases, the NOV24b nucleic acid sequence, located on the q22.12 region of chromosome 21, has 741 of 847 bases (87%) identical to a parchorin mRNA from *oryctolagus cuniculus* gb accno AB035520.1) ($E = 3.2e^{-140}$).

25 A disclosed NOV24b polypeptide (SEQ ID NO:106) encoded by SEQ ID NO:105 has 284 amino acid residues and is presented in Table 24D using the one-letter amino acid code.

5

MAEAAEPEAQLSNHLAEEGPAEGSGEAAARVNGRREDEGEASEPRALGQEHDIITLFVKAGYDGESIGNCPFSQR
LFMILWLKGVIFNVTTVDLKRKPADLQNLAPGTPNPFMTFDGEVKTVDVNKIEEFLEELKAPPRYPKLGTQHP
ESNSAGNDVFAKFSAFIKNTKKDANEIHLKNTLKALRKLDNYLSPLPDEIDAYSTEDTVSRGKFLDGDDEL
TLADCNLLPKLHIIKIVAKYRDDEFPSEMTGIWRVYNNAAYDEFNTPDCADOEIEHAYSADVAKRMK

10

The disclosed NOV24a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 24E.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7592636 dbj BAA94345.1 (AB035520)	parchorin [Oryctolagus cuniculus]	637	436/715 (60%)	475/715 (65%)	e-130
gi 6685319 sp Q9Y696 CLi4_HUMAN	CHLORIDE INTRACELLULAR CHANNEL PROTEIN 4 (INTRACELLULAR CHLORIDE ION CHANNEL PROTEIN P64H1)	253	182/238 (76%)	207/238 (86%)	e-108
gi 7330335 ref NP_039234.1 (NM_013943)	chloride intracellular channel 4; chloride intracellular channel 4 like [<i>Homo sapiens</i>]	253	182/238 (76%)	208/238 (86%)	e-108
gi 7304963 ref NP_038913.1 (NM_013885)	chloride intracellular channel 4 (mitochondrial) [<i>Mus musculus</i>]	253	181/238 (76%)	207/238 (86%)	e-107
gi 4588524 gb AAD26136.1 AF109196_1 (AF109196)	intracellular chloride channel p64H1 [<i>Homo sapiens</i>]	253	180/238 (75%)	205/238 (85%)	e-106

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 24F. In the ClustalW alignment of the NOV24 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 24F. ClustalW Analysis of NOV24

- 1) Novel NOV24a (SEQ ID NO:104)
- 2) Novel NOV24b (SEQ ID NO:106)
- 3) gi|7592636|dbj|BAA94345.1| (AB035520) parchorin [*Oryctolagus cuniculus*] (SEQ ID NO:421)
- 4) gi|6685319|sp|Q9Y696|CL14_HUMAN CHLORIDE INTRACELLULAR CHANNEL PROTEIN 4 (INTRACELLULAR CHLORIDE ION CHANNEL PROTEIN P64H1) (SEQ ID NO:422)
- 5) gi|7330335|ref|NP_039234.1| (NM_013943) chloride intracellular channel 4; chloride intracellular channel 4 like [*Homo sapiens*] (SEQ ID NO:423)
- 6) gi|7304963|ref|NP_038913.1| (NM_013885) chloride intracellular channel 4 (mitochondrial) [*Mus musculus*] (SEQ ID NO:424)
- 7) gi|4588524|gb|AAD26136.1|AF109196_1 (AF109196) intracellular chloride channel p64H1 [*Homo sapiens*] (SEQ ID NO:425)

					10	20	30	40	50	60	
										
25	NOV24a	1			MAEAAEPEGVAPGPQGPPEVPAPLAERPGEPGAAG-----GAEAGPEGSEGAEEAPRGAA						55
	NOV24b	1			-----						1
	gi 7592636 	1			MAETAEPGGAPSPQGPPEGSALLEERPGEPPAGPEASEGAAKAPSGEGAGAAAAGAT						60
	gi 6685319 	1			-----						1
	gi 7330335 	1			-----						1
30	gi 7304963 	1			-----						1
	gi 4588524 	1			-----						1

			70	80	90	100	110	120		
									
35	NOV24a	56	AVKEAGGGGPDRGPEAEARGTRG-----AHGETEAEEGAPEGAEVPQGGGEETSGAQQVE							109
	NOV24b	1	-----							1
	gi 7592636	61	EEASGGRDGEGAGEQPADAGTESGGETPDAKGAQIEAEGAPEGTKAPQLGEEGSSGKGQVE							120
	gi 6685319	1	-----							1
	gi 7330335	1	-----							1
40	gi 7304963	1	-----							1
	qi 4588524	1	-----							1

				130	140	150	160	170	180		
45	NOV24a	110								
	NOV24b	1	GASPGRGAQGEPRGEAQREPEDSAAAPERQEEAEQRPEVPEGSASGEAGDSVDAEGPLGDN							169	
	gi 7592636	121	ESGPDCELRGEAAREAEQGAAAPAPGAQEEAVP-----GDSVDAEG----							163	
	gi 6685319	1	-----								1
	gi 7330335	1	-----								1
50	gi 7304963	1	-----								1
	gi 4588524	1	-----								1

[illegible]

gi|7330335| 152 DEYLNSPLPDEIDENSMEIDKFSIRKFLDGNEMTLADCNLLPKLHIVK----VAKKYR 206
gi|7304963| 152 DEYLNSPLPDEIDENSMEIDKFSIRKFLDGNEMTLADCNLLPKLHIVK----VAKKYR 206
gi|4588524| 152 DEYLNSPLPDEIDENSMEIDKFSIRKFLDGNEMTLADCNLLPKLHIVK----VAKKYR 206

5
670 680 690 700
NOV24a 646 DEFFPSEMTGIWRYLNNAYARDEFTNTCPADCEIEHAYSDVAKRMK- 691
NOV24b 239 DEFFPSEMTGIWRYLNNAYARDEFTNTCPADCEIEHAYSDVAKRMK- 284
10 gi|7592636| 592 DEFFPPEMTGIWRYLNNAYARDEFTNTCPADCEIEHAYSDVAKRMK- 637
gi|6685319| 207 NEDIPKEMTGIWRYLNNAYSRDEFTNTCPSDKEVEIAYSDVAKRLTK 253
gi|7330335| 207 NEDIPKEMTGIWRYLNNAYSRDEFTNTCPSDKEVEIAYSDVAKRLTK 253
gi|7304963| 207 NEDIPKGMTGIWRYLNNAYSRDEFTNTCPSDKEVEIAYSDVAKRLTK 253
gi|4588524| 207 NEDIPKEMTGIWRYLNNAYSRDEFTNTCPSDKEVEIAYSDVAKRLTK 253

15 The gene of invention encodes a homolog of parchorin, a new member of the intracellular chloride channel family. Parchorin was discovered as a 120 kDa phosphoprotein in gastric parietal cells (Urushidani et al., J Membr Biol. 1999 Apr 1;168(3):209-20). Subsequent analysis revealed that this protein had significant homology to the family of intracellular chloride channels, especially in the C terminal domain (Nishizawa et al., J Biol Chem 2000 Apr 14;275(15):11164-73). However, unlike other members of this family,
20 parchorin exists mainly in the cytoplasm and translocated to the plasma membrane upon stimulation of chloride ion efflux. In addition, parchorin shows only two hydrophobic domains relative to the ten to twelve domains seen in other intracellular chloride channels. Tissue expression of parchorin in the rabbit is enhanced in cells that secrete water, like parietal cells,
25 choroid plexus, salivary duct, lacrimal gland, kidney, airway epithelia, and chorioretinal epithelia. It is therefore thought that this protein plays a critical role in these tissues, possibly by modulating chloride ion transport.

Intracellular chloride channels have diverse roles within cells, such as volume regulation, acidification of intracellular vesicles, vectorial transepithelial chloride transport and regulation of cellular excitability (Jentsch et al., Pflugers Arch 1999 May;437(6):783-95). Loss of function mutations affecting three different members of this family lead to three human inherited diseases: myotonia congenita, Dent's disease, and Bartter's syndrome. In addition, a mouse knockout model involving a member of this family has been generated that mimics diabetes insipidus (Matsumura et al., Nat Genet 1999 Jan;21(1):95-8).

35 It is likely, therefore, that the protein of invention participates in physiological functions similar to those of other members of the intracellular chloride channel family, particularly parchorin.

The disclosed NOV24 nucleic acid of the invention encoding a Parchorin-like protein includes the nucleic acids whose sequences are provided in Table 24A or 24C or a fragment
40 thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 24A or 24C while still encoding a

The disclosed NOV24 protein of the invention includes the Parchorin -like protein whose sequence is provided in Table 24B or 24D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 24B or 24D while still encoding a protein that maintains its Adrenal secretory serine protease -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 40 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Parchorin -like protein (NOV24) is a member of a “Parchorin family”. Therefore, the NOV24 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV24 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS.

diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, cancer, trauma, bacterial/viral/parasitic infection, tissue degeneration, and/or other diseases and pathologies.

NOV24 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV24 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV24 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV25

A disclosed NOV25 nucleic acid of 1123 nucleotides (also referred to as CG56457-01) encoding a novel protein phosphatase-like protein is shown in Table 25A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 60-62 and ending with a TGA codon at nucleotides 768-770. The start and stop codons are shown in bold in Table 25A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 25A. NOV25 nucleotide sequence (SEQ ID NO:107).

TCCGGATCGCTTCCCGGGCGGCGAGCTGGGGGTGCACCCGGACCGCCGCCCGGGGATCATGGGCAATGGCA
TGACCAAGGTACTTCTTGGACTCTACCTCGGAAACTTCATTGATGCCAAAGACCTGGATCGCCTGGGCGGAA
ATAAGATCACACACATCATCTCTATCCATGAGTCACCCAGCCTCTGCTGCAGGATATCACCTACCTTCGCA
TCCCGGTGCTGATACCCCTGAGGTACCCATCAAAAAGCACTTCAAAGAATGTATCAACTTCATCCACTGCT
GCCGCCTTAATGGGGGAACTGCCTTGTGCACTGCTTGCAGGCATCTCTCGCAGCACCACGATTGTGACAG
CGTATGTGATGACTGTGACGGGGCTAGGCTGGCGGGACGTGCTTGAAGCCATCAAGGCCACCAGGCCCATCG
CCAACCCCAACCCAGGCTTTAGGCAGCAGCTTGAAGAGTTTGGCTGGGCCAGTTCACAGAAGCTTCGCGGC
AGCTGGAGGAGCGCTTCGGCGAGAGCCCCTTCGCGCAGAGGAGGACTTGCGCGCGCTGCTGCCTCTCTGCA
GGCGCTGTGCGCCAGGCTCCGGGGACTTCGGCCCCGTGCGCCACCACAGCGTCCTCGGCCGCTTCAGAGGGA
CCCTGCAGCGCTGCTGCGCGCATCGCCGCGGAATCACACCGGCCGCTGCGCGCTGCTGCGCGCGTCAAGC
AGACTTTCTCTTGCCTCCCCCGTGTCTGTCCCGCAAGGGCGGCAAGT**AGGATG**CAGTCCAGCCGTGGCTC
CCCACTTCCGACTGGCTCCCTTCGGGGGCTGCTGCGCCTTCACGCCCCCAGGACGGGCCAGAGGCTGG
GGGAGCCCCGCGCGGCTGAACCTGCCTCCCGCGCCCGCCCTGCTCGTCCGCGTCTGCAGTCAGCGTCCC
CAACCTGTGCGTCTCTGTGTCGGGCGGCGCTGCTGCAGCCACCTGGTGCCTTAGTCCTTGGGCTGGGGAG
GGGGCCCCACCTTAAGGGCGGGGGAGGGAGGGAGAGTGGAGGTTTGACGGGCTGGAGGGTATTA
AAGAGACACAGAAGAAGCTGCCTGTCAAAAAAAAAAAAAAAAAA

In a search of public sequence databases, the NOV25 nucleic acid sequence, located on chromosome 20, has 324 of 505 bases (64%) identical to a gb:GENBANK-ID:AF165519|acc:AF165519.1 mRNA from *Homo sapiens* (mitogen-activated protein kinase phosphatase x (MKPX) mRNA, complete cds) ($E = 2.3e^{-31}$).

A disclosed NOV25 polypeptide (SEQ ID NO:108) encoded by SEQ ID NO:107 has 236 amino acid residues and is presented in Table 25B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV25 has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500. Alternatively, NOV25 may also localize to the lysosome (lumen) with a certainty of 0.1805, to the mitochondrial matrix space with a certainty of 0.1000, or to the plasma membrane with a certainty of 0.1000.

Table 25B. Encoded NOV25 protein sequence (SEQ ID NO:108).

MGNGMTKVLPGLYLGNFIDAKDLDRGRNKITHIISIHESQPLLQDITYLRIPVADTPEVPIKKHFKECIN FIHCCRLNGGNCLVHCFAGISRSTTIVTAYVMTVTGLGWRDVLEAIKATRPANPNPGFRQOLEEFGWASSQ KLRRQLEERFGESPFRRDEEDLRALLPLCRRRCRQGPOTSAPSATTASSAASEGTLQRLVPRSPRESHRPLPLL ARVKQTFSCSLPRCLSRKGGK
--

A search of sequence databases reveals that the NOV25 amino acid sequence has 91 of 169 amino acid residues (53%) identical to, and 125 of 169 amino acid residues (73%) similar to, the 184 amino acid residue ptnr:SPTREMBL-ACC:Q9NRW4 protein from *Homo sapiens* (Human) (Mitogen-Activated Protein Kinase Phosphatase X) ($E = 7.3e^{-50}$).

NOV25 is predicted to be expressed in at least brain, testis, exocrine pancreas, adipose, bone, peripheral blood, salivary glands, spinal cord, thyroid. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources and/or RACE sources.

In addition, the sequence is predicted to be expressed in hematopoietic stem cells because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF165519|acc:AF165519.1) a closely related *Homo sapiens* mitogen-activated protein kinase phosphatase x (MKPX) mRNA, complete cds homolog.

The disclosed NOV25 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 25C.

Table 25C. BLAST results for NOV25

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17458347 ref XP_059288.1 (XM_059288)	similar to bA243J16.6 (novel protein with a dual specificity phosphatase, catalytic domain) (H. sapiens) [<i>Homo sapiens</i>]	235	223/236 (94%)	229/236 (96%)	e-124
gi 18104942 ref NP_542178.1 (NM_080611)	dual specificity phosphatase-like 15 [<i>Homo sapiens</i>]	243	216/251 (86%)	222/251 (88%)	e-115

gi 9910432 ref NP_064570.1 (NM_020185)	mitogen-activated protein kinase phosphatase x; homolog of mouse dual specificity phosphatase LMW-DSP2; JNK-stimulating phosphatase 1 [Homo sapiens]	184	91/169 (53%)	125/169 (73%)	4e-53
gi 13183069 gb AAK15038.1 AF237619_1 (AF237619)	dual specificity phosphatase TS-DSP2 [Mus musculus]	184	90/169 (53%)	125/169 (73%)	2e-52
gi 14726046 ref XP_046543.1 (XM_046543)	mitogen-activated protein kinase phosphatase x [Homo sapiens]	184	89/169 (52%)	118/169 (69%)	2e-50

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 25D. In the ClustalW alignment of the NOV25 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 25D. ClustalW Analysis of NOV25

- 1) Novel NOV25 (SEQ ID NO:108)
- 2) gi|17458347|ref|XP_059288.1| (XM_059288) similar to bA243J16.6 (novel protein with a dual specificity phosphatase, catalytic domain) (H. sapiens) [Homo sapiens] (SEQ ID NO:426)
- 3) gi|18104942|ref|NP_542178.1| (NM_080611) dual specificity phosphatase-like 15 [Homo sapiens] (SEQ ID NO:427)
- 4) gi|9910432|ref|NP_064570.1| (NM_020185) mitogen-activated protein kinase phosphatase x; homolog of mouse dual specificity phosphatase LMW-DSP2; JNK-stimulating phosphatase 1 [Homo sapiens] (SEQ ID NO:428)
- 5) gi|13183069|gb|AAK15038.1|AF237619_1 (AF237619) dual specificity phosphatase TS-DSP2 [Mus musculus] (SEQ ID NO:429)
- 6) gi|14726046|ref|XP_046543.1| (XM_046543) mitogen-activated protein kinase phosphatase x [Homo sapiens] (SEQ ID NO:430)

			10	20	30	40	50	60							
NOV25	1	MGNGMTKVL	PGLYL	GNFI	-----	DAKDL	DRLGR	NKITHI	SIHES	POPLL	45				
gi 17458347	1	MGNGMTKVL	PGLYL	GNFI	-----	DAKDL	DOLGR	NKITHI	SIHES	POPLL	45				
gi 18104942	1	MGNGMTKVL	PGLYL	GNFIGH	PASQIGSSILFLS	DAKDL	DOLGR	NKITHI	SIHES	POPLL	60				
gi 9910432	1	MGNGMKNIL	PGLYL	GNFK	-----	DARDAE	QLSKNK	VTHILSV	HDSAR	PML	45				
gi 13183069	1	MGSGMSQIL	PGLYL	GNFK	-----	DARDAE	QLSRNK	VTHILSV	HDTAR	PML	45				
gi 14726046	1	MGNGMKNIL	PGLYL	GNFK	-----	DARDAG	QLSRNK	VTHILSV	HDXXXXXX		45				
			70	80	90	100	110	120							
NOV25	46	QDITYLRL	IPVAD	TPEVP	IKKH	FKEC	INF	IHC	CRLNG	GNCLVHCF	FAGISR	STTIV	TAYV	MT	105
gi 17458347	46	QDITYLRL	IPVAD	TPEVP	IKKH	FKEC	INF	IHC	CRLNG	GNCLVHCF	FAGISR	STTIV	TAYV	MT	105
gi 18104942	61	QDITYLRL	IPVAD	TPEVP	IKKH	FKEC	INF	IHC	CRLNG	GNCLVHC	-----	TTTIV	TAYV	MT	113

	gi 9910432	46	EGVKYLCIPAAADSPSONLIRHFKESIKFIHBCRLRGESCLVHCLACVSRSVTLVIAYIMT	105
	gi 13183069	46	EGVKYLCIPAAADSPSONLIRHFKESIKFIHBCRLRGESCLVHCLACVSRSVTLVIAYIMT	105
	gi 14726046	46	XGVKYLCLPAADSPSONLIRHFKESIKFIHBCRLRGESCLVHCLACVSRSVTLVIAYIMT	105
5			130 140 150 160 170 180	
	NOV25	106	VTGLGWRDVLEAIKATRPIANPNPGFRQOLEEFGWASSQKLRRROLEERFGESPFDEEDL	165
	gi 17458347	106	VTGLGWRDVLEAIKATRPIANPNPGFRQOLEEFGWASSQKLRRROLEERFGESPFDEEDL	165
	gi 18104942	114	VTGLGWRDVLEAIKATRPIANPNPGFRQOLEEFGWASSQKLRRROLEERFGESPFDEEDL	173
10	gi 9910432	106	VTDFGWEDALHTVRAGRSCANPNVGFQROLQEFKHEVHQYRQWLKEEYGESPLQDAEBA	165
	gi 13183069	106	VTDFGWEDALHTVRAGRSCANPNVGFQROLQEFKHEVHQYRQWLKEEYGESPLQDAEBA	165
	gi 14726046	106	VTDFGWEDALHTVRAGRSCANPNVGFQROLQEFKHEVHQYRQWLKEEYGESPLQDAEBA	165
			190 200 210 220 230 240	
15	NOV25	166	RALLPLCRRCRQGPSTAPSATTASSAASECTLQRLVPRSPRESHRPLPLLARVKQTESC	225
	gi 17458347	166	RALLPLCKRCRQGSATSASSAG-PHSAASECTVQRLVPRTPREAHRPLPLLARVKQTESC	224
	gi 18104942	174	RALLPLCKRCRQGSATSASSAG-PHSAASECTVQRLVPRTPREAHRPLPLLARVKQTESC	232
	gi 9910432	166	KNIL-----AAPGILK-----FWA	179
20	gi 13183069	166	KNIL-----AAPGILK-----FWA	179
	gi 14726046	166	KNIL-----AAPGIMK-----FWA	179
			250	
25	NOV25	226	LPRCLSRKGGK	236
	gi 17458347	225	LPRCLSRKGGK	235
	gi 18104942	233	LPRCLSRKGGK	243
	gi 9910432	180	FLRRL-----	184
	gi 13183069	180	FLRRL-----	184
30	gi 14726046	180	FLRRL-----	184

Tables 25E-H list the domain descriptions from DOMAIN analysis results against NOV25. This indicates that the NOV25 sequence has properties similar to those of other proteins known to contain this domain.

Table 25E Domain Analysis of NOV25

gnl|Smart|smart00195, DSPc, Dual specificity phosphatase, catalytic domain (SEQ ID NO:816)
 CD-Length = 139 residues, 97.8% aligned
 Score = 139 bits (349), Expect = 2e-34

	NOV25:	4	GMTKVLPGLYLGNFIDAKDLRLGRNKITHIIS-IHESQPILLQDITYLRIPVADTPEVP	62
40	Sbjct:	1	GPSEILPHLYLGSYSASNALLKLGITHVINVTVEVPNSNKSGLYLGIPVDDNTETK	60
	NOV25:	63	IKKHFKECINFHCCRLNGGNCLVHCFAGISRSTTIVTAYVMTVTGLGWRDVLEAIKATR	122
45	Sbjct:	61	ISPYLPEAVEFIEDAEKKGKVLVHCQAGVSRSATLI IAYLMKYRNMSLNDAIDFVKERR	120
	NOV25:	123	PIANPNPGFRQOLEEF	138
	Sbjct:	121	PIISENFGFLRQLIEY	136

1950年12月，中共中央、国务院发布《关于整顿和统一全国图书出版业的决定》，规定出版业实行“归口管理”，即出版业归文化系统管理，其他系统不得出版图书。这一决定旨在加强出版业的统一管理，提高出版质量，防止重复出版和浪费资源。同时，还规定出版业实行“计划管理”，即出版计划由国家统一制定，各出版单位必须严格执行。这一系列措施为出版业的健康发展奠定了基础。

5

10

gnl|Smart|smart00404, PTPc_motif, Protein tyrosine phosphatase,
catalytic domain motif (SEQ ID NO:818)
CD-Length = 105 residues, 53.3% aligned
Score = 41.2 bits (95), Expect = 7e-05

15

gnl|Pfam|pfam00102, Y_phosphatase, Protein-tyrosine phosphatase (SEQ
ID NO:819)
CD-Length = 235 residues, 31.9% aligned
Score = 38.5 bits (88), Expect = 4e-04

20

25

30

inherited and sporadic cancers (Eng, Recent Prog Horm Res 1999;54:441-52; discussion 453). In *Drosophila*, members of the DSP family, such as puckered, have important roles in development (Martin-Blanco et al., Genes Dev 1998 Feb 15;12(4):557-70). The crystal structure of one member of the DSP family has been elucidated (Yuvaniyama et al., Science 1996 May 31;272(5266):1328-31) and this family has been successfully targeted for small molecule drug development (Ducruet et al., Bioorg Med Chem 2000 Jun;8(6):1451-66). In addition, overexpression of a DSP has been demonstrated to be a potential therapy for cardiac hypertrophy (Bueno et al., Circ Res 2001 Jan 19;88(1):88-96). The gene of invention has greatest homology to a DSP identified in hematopoietic stem/progenitor cells from a patient with myelodysplastic syndromes. It shows the presence of a distinct domain present in all DSPs, which qualifies it as a *bona fide* member of this family. Its localization is predicted to be cytoplasmic, which makes it a good candidate to interact with members of the signal transduction cascade governing the cell cycle.

The disclosed NOV25 nucleic acid of the invention encoding a Protein phosphatase -like protein includes the nucleic acid whose sequence is provided in Table 25A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 25A while still encoding a protein that maintains its Protein phosphatase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36 percent of the bases may be so changed.

The disclosed NOV25 protein of the invention includes the Protein phosphatase -like protein whose sequence is provided in Table 25B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 25B while still encoding a protein that maintains its Protein phosphatase -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 48 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Protein phosphatase -like protein (NOV25) is a member of a "Protein phosphatase family". Therefore, the NOV25 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV25 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in for example Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, psychiatric disorders, metabolic disorders, fertility, hypogonadism, xerostomia, hyperthyroidism, hypothyroidism, cancer, trauma, tissue degeneration, viral/bacterial/parasitic infections, and/or other diseases and pathologies.

NOV25 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV25 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV25 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV26

NOV26 includes two novel GAGE-7-like proteins disclosed below. The disclosed sequences have been named NOV26a and NOV26b.

NOV26a

A disclosed NOV26a nucleic acid of 550 nucleotides (also referred to as CG56461-01) encoding a novel GAGE-7-like protein is shown in Table 26A. An open reading frame was

identified beginning with a ATG initiation codon at nucleotides 67-69 and ending with a TAA codon at nucleotides 400-402. The start and stop codons are shown in bold in Table 26A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 26A. NOV26a nucleotide sequence (SEQ ID NO:109).

GTTCCTGCTGTCTGGACTTTTCTGTCCCACTGAGACGCAGCTGTATTCTGTTTGAGCTGTGAAAT**ATG**ATT
TGGCGAGGAAGATCAACATATAGGCCTAGGCCGAGGAGAAGTGACACCTCCTGAGCTGATTGGGCCTATG
CTGGAGCCCGGTGATGAGGAGCCTCAGCAAGAGGAACCCCAACTGAAAGTCGGGATCCTGCACCTGGTCAG
GAGAGAGAAGAAGATCAGGGTGCAGCTGAGACTCAAGTGCCCTGACCTGGAAGCTGATCTCCAGGAGCTGTCT
CAGTCAAAGACTGGGGGTGAATGTGGAAATGGTCTGATGACCAGGGGAAGATTCTGCCAAAATCAGAACAA
TTTAAATGCCAGAAGGAGGTGACAGGCAACCACAGGTT**TAA**ATGAAGACAAGCTGAAACAACACAAAAC**TG**
TTTTTATCTAAGATATTGACTTAAAAATATCGAAATAAACTTTTGCAGCTTTCTCCAAAAA
AAACCCGC

In a search of public sequence databases, the NOV26a nucleic acid sequence, located on the X chromosome, has 293 of 360 bases (81%) identical to a gb:GENBANK-ID:AF251237|acc:AF251237.1 mRNA from *Homo sapiens* (XAGE-1 mRNA, complete cds) ($E = 3.6e^{-46}$).

A disclosed NOV26a polypeptide (SEQ ID NO:110) encoded by SEQ ID NO:109 has 111 amino acid residues and is presented in Table 26B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV26a has no signal peptide and is likely to be localized to the mitochondrial matrix space with a certainty of 0.4462. Alternatively, NOV26A may also localize to the nucleus with a certainty of 0.3000, to the mitochondrial inner membrane with a certainty of 0.1347, or to the mitochondrial intermembrane space with a certainty of 0.1347.

Table 26B. Encoded NOV26a protein sequence (SEQ ID NO:110).

MIWRGRSTYRPRPRRSVPPPELIGPMLFPGDEEPQQEPPPTESRDPAPGQEREEDQGAETQVPDLEADLQE
LSQSKTGGECCNGPDDQKILPKSEQFKMPEGGRQPQV

A search of sequence databases reveals that the NOV26a amino acid sequence has 60 of 115 amino acid residues (52%) identical to, and 72 of 115 amino acid residues (62%) similar to, the 116 amino acid residue ptnr:SPTREMBL-ACC:Q9UEU5 protein from *Homo sapiens* (Human) (GAGE-7) ($E = 1.4e^{-23}$).

NOV26a is predicted to be expressed in at least placenta. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, and/or RACE sources.

NOV26b

In the present invention, the target sequence identified previously, NOV26a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV26b. This is 100% identical to the previously identified sequence (NOV26a).

A disclosed NOV26b nucleic acid of 494 nucleotides (also referred to as CG56461-02) encoding a novel GAGE-7-like protein is shown in Table 26C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 67-69 and ending with a TAA codon at nucleotides 400-402. The start and stop codons are shown in bold in Table 26C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 26C. NOV26b nucleotide sequence (SEQ ID NO:111).

GTTCCTGCTGTCTGGACTTTTCTGTCCCACTGAGACGCAGCTGTATTCTGTTTGCACTGTGAAATATGATT
TGGCGAGGAAGATCAACATATAGGCCTAGGCCGAGGAGAAGTGTACCACCTCCTGAGCTGATTGGGCCTATG
CTGGAGCCCGGTGATGAGGAGCCTCAGCAAGAGGAACCACCAACTGAAAGTCGGGATCCTGCACCTGGTCAG
GAGAGAGAAGAAGATCAGGGTGCAGCTGAGACTCAAGTGCCCTGACCTGGAAGCTGATCTCCAGGAGCTGTCT
CAGTCAAAGACTGGGGGTGAATGTGAAATGGTCCTGATGACCAGGGGAAGATTCTGCCAAAATCAGAACAA
TTTAAATGCCAGAAGGAGGTGACAGGCAACCACAGGTTTAAATGAAGACAAGCTGAAACAACACAAAACCTG
TTTTTATCTAAGATATTGACTTAAAAATATCAAATAAACTTTTCAGCTTTCTCCAAAAA

In a search of public sequence databases, the NOV26b nucleic acid sequence, located on the X chromosome, has 346 of 426 bases (81%) identical to a gb:GENBANK-ID:HSL185E6A|acc:Z68274.1 mRNA from *Homo sapiens* (Human DNA sequence from cosmid L129H7, Huntington's Disease Region, chromosome 4p16.3 contains Pseudogene and CpG island) ($E = 5.7e^{-53}$).

The disclosed NOV26b polypeptide (SEQ ID NO:112) encoded by SEQ ID NO:111 has 111 amino acid residues and is presented in Table 26D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV26b has no signal peptide and is likely to be localized to the mitochondrial matrix space with a certainty of 0.4462. Alternatively, NOV26b may also localize to the nucleus with a certainty of 0.3000, to the mitochondrial inner membrane with a certainty of 0.1347, or to the mitochondrial intermembrane space with a certainty of 0.1347.

Table 26D. Encoded NOV26b protein sequence (SEQ ID NO:112).

MIWRGRSTYRPRPRRSVPPPELIGPMLEPGDEEPQQEPPPTESRDPAPGQEREEDQGAETQVPDLEADLQE
LSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPQV

A search of sequence databases reveals that the NOV26b amino acid sequence has 60 of 115 amino acid residues (52%) identical to, and 72 of 115 amino acid residues (62%) similar to, the 116 amino acid residue ptnr:SPTREMBL-ACC:Q9UEU5 protein from *Homo sapiens* (Human) (GAGE-7) ($E = 1.4e^{-23}$).

NOV26b is predicted to be expressed in at least the following tissues: Placenta, Whole Organism. .

The disclosed NOV26a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 26E.

Table 26E. BLAST results for NOV26a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17486397 ref XP_060048.1 (XM_060048)	similar to G antigen 3 (H. sapiens) [<i>Homo sapiens</i>]	137	84/84 (100%)	84/84 (100%)	2e-33
gi 18027836 gb AAL55879.1 AF318372_1 (AF318372)	unknown [<i>Homo sapiens</i>]	111	110/111 (99%)	110/111 (99%)	4e-33
gi 18157212 emb CAC83008.1 (AJ318881)	XAGE-3 protein [<i>Homo sapiens</i>]	111	111/111 (100%)	111/111 (100%)	2e-29

gi 17486394 ref XP_066835.1 (XM_066835)	similar to G antigen B1; prostate associated gene 1 (H. sapiens) [Homo sapiens]	185	64/78 (82%)	69/78 (88%)	2e-26
gi 14765261 ref XP_032309.1 (XM_032309)	hypothetical protein XP_032309 [Homo sapiens]	111	80/111 (72%)	93/111 (83%)	4e-26

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 26F. In the ClustalW alignment of the NOV26 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 26F. ClustalW Analysis of NOV26

- 1) Novel NOV26a (SEQ ID NO:110)
- 2) Novel NOV26b (SEQ ID NO:112)
- 3) gi|17486397|ref|XP_060048.1| (XM_060048) similar to G antigen 3 (H. sapiens) [Homo sapiens] (SEQ ID NO:431)
- 4) gi|18027836|gb|AAL55879.1|AF318372.1 (AF318372) unknown [Homo sapiens] (SEQ ID NO:131) gi|18157212|emb|CAC83008.1| (AJ318881) (SEQ ID NO:432)
- 6) gi|18157212|emb|CAC83008.1| (AJ318881) XAGE-3 protein [Homo sapiens] (SEQ ID NO:433)
- 7) gi|17486394|ref|XP_066835.1| (XM_066835) similar to G antigen B1; prostate associated gene 1 (H. sapiens) [Homo sapiens] (SEQ ID NO:434)
- 8) gi|14765261|ref|XP_032309.1| (XM_032309) hypothetical protein XP_032309 [Homo sapiens] (SEQ ID NO:435)

			10	20	30	40	50	60	
NOV26a	1	MIWR	GRSTYRP			RPR	RSVP		18
NOV26b	1	MIWR	GRSTYRP			RPR	RSVP		18
gi 17486397	1	MKNKYVESCWRNVERL				MI RN IYVFCIFVLLKCIDNKT			37
gi 18027836	1	MIWR	GRSTYRH			RPR	RSVP		18
gi 18157212	1	MIWR	GRSTYRP			RPR	RSVP		18
gi 17486394	1	MSLIWATKATTKYTTIDEMCASSADISVCFAQSCGRKRHDFFLAN				YDTAVPETPGTP			58
gi 14765261	1	MSWR	GRSTYRP			RPR	RSVP		18
			70	80	90	100	110	120	
NOV26a	18		PEELIGPML	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA					68
NOV26b	18		PEELIGPML	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA					68
gi 17486397	37	--FYLHTHTHTHTHFWFQ	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA						94
gi 18027836	18		PEELIGPML	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA					68
gi 18157212	18		PEELIGPML	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA					68
gi 17486394	59	NRGQIPDSSRDIAKAGMSWT	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA						118
gi 14765261	18		PEELIGPML	EPGDEEPQQUEPPTESRDPAPGQEREEDQGAETQVPDLEA					68
			130	140	150	160	170	180	
NOV26a	69		DLQELSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPV						111
NOV26b	69		DLQELSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPV						111
gi 17486397	95		DLQELSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPV						137

```

5
gi|18027836| 69 DLQELSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPQV----- 111
gi|18157212| 69 DLQELSQSKTGGECGNGPDDQGKILPKSEQFKMPEGGDRQPQV----- 111
gi|17486394| 119 DLQELSQSKTGDCCGDSPDVQKILPKSEQFKMPEGGFLCSRCILSKLYTMFAIYFKPLL 178
gi|14765261| 69 DLQELCQTKTGDCCEGGTDVKGKILPKAEHFKMPEAGEGKSOV----- 111

10
NOV26a      111 ----- 111
NOV26b      111 ----- 111
gi|17486397| 137 ----- 137
gi|18027836| 111 ----- 111
gi|18157212| 111 ----- 111
gi|17486394| 179 IDALDLM 185
15 gi|14765261| 111 ----- 111

```

The gene of invention is a member of GAGE family of proteins. It belongs to the broad family of GAGE/MAGE/PAGE genes that are expressed in various human cancers. Many human tumors express antigens that are recognized *in vitro* by cytolytic T lymphocytes (CTLs) derived from the tumor-bearing patient. The MAGE(melanoma-specific antigen), PAGE (Prostate cancer antigen) and GAGE (G antigen) gene family members encode such antigens. Therefore these antigens can serve as therapeutic targets in cancer.

The LNCaP progression model of human prostate cancer consists of lineage-related sublines that differ in their androgen sensitivity and metastatic potential. A differential display polymerase chain reaction was employed by Chen ME, et al. (J Biol Chem 1998 Jul 10;273(28):17618-25) to evaluate mRNA expression differences between the LNCaP sublines in order to define the differences in gene expression between the androgen-sensitive, nontumorigenic LNCaP cell line and the androgen-insensitive, metastatic LNCaP sublines, C4-2 and C4-2B. An amplicon, BG16.21, was isolated that showed increased expression in the androgen-independent and metastatic LNCaP sublines, C4-2 and C4-2B. Hybridization screening of a lambda gt11 expression library with BG16.21 revealed two transcripts, both homologous to BG16.21 at the 3' end. A GenBank™ data base search using the GCG Wisconsin software package revealed the shorter approximately 600-bp transcript (designated GAGE-7) to be a new member of the GAGE family. The second approximately 700-bp transcript was a novel gene (designated PAGE-1, "prostate associated gene") with only 45% homology to GAGE gene family members. RNA blot analysis demonstrated that GAGE-7 mRNA was expressed at equal levels in all lineage related prostate cancer cell sublines, while PAGE-1 mRNA levels were elevated 5-fold in C4-2 and C4-2B as compared with LNCaP cells. Neither GAGE-7 nor PAGE-1 demonstrated any regulation by androgens in the prostate cancer cell lines used in this study. PAGE-1 and GAGE-7 expression was found to be restricted to testes (high) and placenta (low) on human multiple tissue Northern blots. As GAGE/MAGE antigens were reported previously to be targets for tumor-specific cytotoxic

lymphocytes in melanoma, these results suggest that PAGE-1 and GAGE-7 may be related to prostate cancer progression and may serve as potential targets for novel therapies.

The GAGE-1 gene was identified previously as a gene that codes for an antigenic peptide, YRPRPRRY, which was presented on a human melanoma by HLA-Cw6 molecules and recognized by a clone of CTLs derived from the patient bearing the tumor. By screening a cDNA library from this melanoma, De Backer O, et al. (*Cancer Res* 1999 Jul 1;59(13):3157-65) identified five additional, closely related genes named GAGE-2-6. We report here that further screening of this library led to the identification of two more genes, GAGE-7B and -8. GAGE-1, -2, and -8 code for peptide YRPRPRRY. Using another antitumor CTL clone isolated from the same melanoma patient, they identified antigenic peptide, YYWPRPRRY, which is encoded by GAGE-3, -4, -5, -6, and -7B and which is presented by HLA-A29 molecules. Genomic cloning of GAGE-7B showed that it is composed of five exons. Sequence alignment showed that an additional exon, which is present only in the mRNA of GAGE-1, has been disrupted in gene GAGE-7B by the insertion of a long interspersed repeated element retroposon. These GAGE genes are located in the p11.2-p11.4 region of chromosome X. They are not expressed in normal tissues, except in testis, but a large proportion of tumors of various histological origins express at least one of these genes. Treatment of normal and tumor cultured cells with a demethylating agent, azadeoxycytidine, resulted in the transcriptional activation of GAGE genes, suggesting that their expression in tumors results from a demethylation process.

The disclosed NOV26 nucleic acid of the invention encoding a GAGE-7 -like protein includes the nucleic acid whose sequence is provided in Table 26A, 26C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 26A or 26C while still encoding a protein that maintains its GAGE-7 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense

binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19 percent of the bases may be so changed.

5 The disclosed NOV26 protein of the invention includes the GAGE-7 -like protein whose sequence is provided in Table 26B or 26D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 26B or 26D while still encoding a protein that maintains its GAGE-7 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 28 percent of the residues may be so changed.

10 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this GAGE-7 -like protein (NOV26) is a member of a "GAGE-7 family". Therefore, the NOV26 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

20 The NOV26 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in fertility disorders, cancer, trauma, tissue degeneration, viral/bacterial/parasitic infections, and/or other diseases and pathologies.

NOV26 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV26 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV26 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV27

NOV27 includes three novel sodium - glucose cotransporter -like proteins disclosed below. The disclosed sequences have been named NOV27a, NOV27b, and NOV27c.

NOV27a

A disclosed NOV27a nucleic acid of 1914 nucleotides (also referred to as CG56645-01) encoding a novel sodium - glucose cotransporter-like protein is shown in Table 27A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 51-53 and ending with a TGA codon at nucleotides 1839-1841. The start and stop codons are shown in bold in Table 27A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 27A. NOV27a nucleotide sequence (SEQ ID NO:113).

TTGCCCCCTCAGTCCCTCGGGCTCATACCTAGTGCCTGCGGCAGGACAGCCATGGCCGCCAACTCCACCAGCG
ACCTCCCACTCCCGGACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGATTTTGGCTCTGAACG
TGGCCGTGGGCATATGGTCTCTTGTGCGGGCCAGTAGGAACACGGTGAATGGCTACTTCTGGCAGGCCGGG
ACATGACGTGGTGGCCGATGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGCCTCTTCATTGGACTGG
CGGGCTCAGGCGCGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAATGCCACGTACGTGCTGCTGGCAC
TGGCATGGGTGTTTCGTGCCCATCTACATCTCCTCAGAGATCGTCACCTTACCTGAGTACATTGAGAAGCGCT
ACGGGGGCCAGCGGATCCGCATGTACCTGTCTGTCTCCTGTCCCTGCTACTGTCTGTCTTCACCAAGATATCGC
TGGACCTGTACGCGGGGGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACCTTCTACCTCTCCACCATCTCA
CGCTCGGCATCACAGCCCTGTACACCATCGCAGGGGGCTGGCTGCTGTAATCTACACGGACGCCCTGCAGA
CGCTCATCATGGTGGTGGGGCTGTTCATCCTGACAATCAAAGCTTTTGACCAGATCGGTGGTTACGGGCGAGC
TGGAGGCAGCCTACGCCCAGGCCATTCCTCCAGGACCATGCGCAACACCACCTGCCACCTGCCACGTACAG
ACGCCATGCACATGTTTCGAGACCCCCACACAGGGGACCTGCCGTGGACCGGGATGACCTTTGGCCTGACCA
TCATGGCCACCTGGTACTGGTGACCGACCGAGGTGATCGTGACCGATCACTGTCAGCCCGGACCTGAACC
ATGCCAAGGCGGGCTCCATCTCGCCAGCTACCTCAAGATGCTCCCATGGGCTGATCATAATGCCGGGCA
TGATCAGCCCGCATTTGTTCCAGATGATGTGGGCTGCGTGGTGCCGTCGAGTGCTGCGGGCCTGCGGGG
CCGAGGTCGGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATGGAACGATGCCCATCGGTCTGCGGGGGC
TGATGATCGCAGTGATGCTGGCGGGCTCATGTGCTGCTGACTCCATCTTCAACAGCAGCAGCACCTCT
TCACTATGCACATCTGGAGGCGGCTGCGTCCCGCTCCGCGCAGCGGGAGCTCCTGCTGGTGGGACGGCTGG
TCATAGTGGCACTCATCGGCGTGAGTGTGGCCTGGATCCCCGTCCTGCAGGACTCCAACAGCGGGCAACTCT
TCATCTACATGCAGTCAGTGACCGAGCTCCCTGGCCCCCAGTGAAGTGCAGTCTTTGTCTGGGCGTCTTCT
GGCGACGTGCCAACGAGCAGGGGGCTTCTGGGGCTGATAGCAGGGCTGGTGGTGGGGGCCACGAGGCTGG
TCCTGGAATTCCTGAACCCAGCCCCACCGTGGGAGAGCCAGACACGCGGCCAGCGCTCCTGGGAGCATCC
ACTACCTGCACCTTCGCTGTGCCCCCTTTGCACTCAGTGGTGTGTTGTGGTGGCTGGAAGCCTGCTGACCC
CACCCCCACAGAGTGTCCAGATTGAGAACCTTACCTGGTGGACCTGGCTCAGGATGTGCCCTTGGGAACCTA
AAGCAGGTGATGGCCAAACACCCAGAAACACGCCTTCTGGGCCGCTGTCTGTGGCTTCAATGCCATCTCC
TCATGTGTGTAACATATTCTTTTATGCCTACTTCGCCTGACACTGCCATCCTGGACAGAAAGGCAGGAGCT
CTGAGTCCTCAGGTCCACCATTTCCTCATGGGGATCCCCA

In a search of public sequence databases, the NOV27a nucleic acid sequence, located on chromosome 17, has 1598 of 1838 bases (86%) identical to a gb:GENBANK-
ID:OCU08813|acc:U08813.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus*
Na⁺/glucose cotransporter-related protein mRNA, complete cds) (E = 2.6e⁻³⁰⁹).

A disclosed NOV27a polypeptide (SEQ ID NO:114) encoded by SEQ ID NO:113 has 596 amino acid residues and is presented in Table 27B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV27A has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8200. Alternatively, NOV27a may also localize to the endoplasmic reticulum (membrane) with a certainty of 0.6850, to the Golgi body with a certainty of 0.4600, or to the enoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV27A is between positions 42 and 43: CRA-SR.

Table 27B. Encoded NOV27a protein sequence (SEQ ID NO:114).

```
MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWWPIGASLFASSEG
SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFPYIISSEIVTLPEYIQKRYGGQRIRMYLSVLSLLL
SVFTKISLDLYAGALFVHICLGWNFYLTILTGLITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIIKAFD
QIGGYGQLEAAYAQAIPSRITANTTCHLPRTDAMHMFDPHTGDLPTWGTGTFGLT IMATWYWC TDQVIVQRS
LSARDLNHAKAGSILASYLKMLPMGLIIMPGMTSRALFPDDVGCVPSECLRACGAEVGCSNIAYPKLVMEI
MPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRSGERELLVGRLLVIVALIGVSAWIPVLQ
DSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTR
PAVLGSIHYLHFAVALFALSGAVVAGSLTTPPPQSVQIENLTWWTLAQDVPLGKAGDGQTPQKHAFWARV
CGFNAILLMCVNIFFYAYFA
```

A search of sequence databases reveals that the NOV27a amino acid sequence has 531 of 596 amino acid residues (89%) identical to, and 559 of 596 amino acid residues (93%)

5 similar to, the 597 amino acid residue ptrn:SPTREMBL-ACC:Q28610 protein from *Oryctolagus cuniculus* (Rabbit) (Na⁺/Glucose Cotransporter-Related Protein) ($E = 1.1 \times 10^{-289}$).

NOV27a is predicted to be expressed in at least heart and kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, and/or RACE sources.

10 In addition, the sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:OCU08813|acc:U08813.1) a closely related *Oryctolagus cuniculus* Na⁺/glucose cotransporter-related protein mRNA, complete cds homolog.

NOV27b

15 In the present invention, the target sequence identified previously, NOV27a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a

20 suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in

25 PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea,

uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV27b. This differs from the previously identified sequence (NOV27a) in having 16 extra internal amino acids.

A disclosed NOV27b nucleic acid of 1912 nucleotides (also referred to as CG56645-02 encoding a novel sodium - glucose cotransporter-like protein is shown in Table 27C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 35-37 and ending with a TGA codon at nucleotides 1871-1873. The start and stop codons are shown in bold in Table 27C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 27C. NOV27b nucleotide sequence (SEQ ID NO:115).

CGGGCTCATACCTAGTGCCTGCGGCAGGACAGCCATGGCCGCCAACTCCACCAGCGACCTCCACACTCCCGG
GACGCAGCTGAGCGTGGCTGACATCATCGTCATCACTGTGTATTTGCTCTGAACGTGGCCGTGGGCATATG
GTCCTCTGTGTCGGGCCAGTAGGAACACGGTGAATGGCTACTTCCCTGGCAGGCCGGACATGACGTGGTGGCC
GATTGGAGCCTCCCTCTTCGCCAGCAGCGAGGGCTCTGGCCTCTTCATTGGACTGGCGGGCTCAGGCGCGGC
AGGAGGTCTGGCCGTGGCAGGCTTCGAGTGGAAATGCCACGTACGTGCTGCTGGCACTGGCATGGGTGTTCTGT
GCCATCTACATCTCCTCAGAGATCGTCACCTTACCTGAGTACATTGAGAAGCGCTACGGGGGCCAGCGGAT
CCGCATGTACCTGTCTGTCTGTCCCTGCTACTGTCTGTCTTCCACCAAGATATCGTGGACCTGTACGCGGG
GGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACCTTACCTCTCCACCATCCTCACGCTCGGCATCACAGC
CCTGTACACCATCGCAGGGGGCCTGGCTGCTGTAATCTACACGGACGCCCTGCAGACGCTCATCATGGTGGT
GGGGGCTGTATCCTGACAATCAAAGCTTTTGACCAGATCGGTGGTTACGGGCAGCTGGAGGCAGCCTACGC
CCAGGCCATTCCCTCCAGGACCATGCCCAACACCACTGCCACCTGCCACGTACAGACGCCATGCACATGTT
TCGAGACCCCCACACAGGGGACCTGCCGTGGACCGGGATGACCTTGGCCTGACCATCATGGCCACCTGGTA
CTGGTGACCGACCGAGGTATCGTGCAGCGATCACTGTGAGCCCGGGACCTGAACCATGCCAAGCGGGCTC
CATCCTGGCCAGCTACCTCAAGATGCTCCCATGGGCCTGATCATCATGCGGGCATGATCAGCCGCGCAT
GTTCCAGGTGCTCATGTCTATGAGGAGAGACCAAGTGTCCGTCTCTCGAACAGATGATGTGGGCTGCGT
GGTGCCGTCCGAGTGCCTGCGGGCCTGCGGGCCGAGGTGCGCTGCTCCAACTACGCTACCCCAAGCTGGT
CATGGAAGTATGATGCCATCGGTCTGCGGGGGCTGATGATCGCAGTATGCTGGCGGCGCTCATGTGCTCGCT
GACCTCCATCTTCAACAGCAGCAGCACCTCTTCACTATGGACATCTGGAGCGGCTGCGTCCCGCTCCGG
CGAGCGGGAGCTCCTGCTGGTGGGACGGCTGGTCATAGTGGCACTCATCGGCGTGAGTGTGGCCTGGATCCC
CGTCTGCAGGACTCCAACAGCGGGCAACTCTTCACTACATGCAGTCAGTGACAGCTCCCTGGCCCCACC
AGTACTGCAGTCTTTGTCTGGGCTCTTCTGGCGACGTGCCAACGAGCAGGGGGCTTCTGGGCGCTGAT
AGCAGGCTGGTGGTGGGGCCAGAGGCTGGTCTGGAATTCCTGAACCCAGCCCCACCGTGGGAGAGCC
AGACACGCGGCCAGCCGTCTGGGGAGCATCCACTACCTGCACCTCGCTGTGCGCCTCTTGCACCTCAGTGG
TGCTGTTGTGGTGGCTGGAAGCCTGTGACCCACCCACAGAGTGTCCAGATTGAGAACCTTACCTGGTG
GACCTGGCTCAGGATGTGCCCTTGGGAACATAAGCAGGTGATGGCCAAACACTCCAGAAACACGCCCTTCTG
GGCCCGTGTCTGTGGCTTCAATGCCATCCTCCTCATGTGTGTCAACATATTCTTTATGCTACTTCCGCTG
ACACTGCCATCCTGGACAGAAAGGCAGGAGCTCTGAGTCC

In a search of public sequence databases, the NOV27b nucleic acid sequence, located on chromosome 17, has 903 of 1017 bases (88%) identical to a gb:GENBANK-ID:OCU08813|acc:U08813.1 mRNA from *Oryctolagus cuniculus* (*Oryctolagus cuniculus* Na⁺/glucose cotransporter-related protein mRNA, complete cds) (E = 4.4e⁻¹⁷⁶).

The disclosed NOV27b polypeptide (SEQ ID NO:116) encoded by SEQ ID NO:115 has 612 amino acid residues and is presented in Table 27D using the one-letter amino acid

code. Signal P, Psort and/or Hydropathy results predict that NOV27b has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8200. Alternatively, NOV27b may also localize to the endoplasmic reticulum (membrane) with a certainty of 0.6850, to the Golgi body with a certainty of 0.4600, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV27B is between positions 42 and 43: CRA-SR.

Table 27D. Encoded NOV27b protein sequence (SEQ ID NO:116).

MAANSTSDLHTPGTQLSVADIIVITVPALNVAVGWSSCRASRNTVNGYFLAGRDMTWPIGASLFASSEG
SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSEIVTLPEYIQKRYGGQRIRMYLSVLSLLL
SVFTKISLDLYAGALFVHICLGWNFYLSLTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIFKAFD
QIGGYGQLEAAYAQAIPSRITANTTCHLPRTDAMHMFDPHTGDLPTGTMFTGLTIMATWYWCDDQVIVQRS
LSARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPGAHVYEEERHQVSVSRDQVGVVPECLRACGA
EVGCSNIAYPKLVMEMLPIGLRGLMIAVMLAALMSSSLTSIFNSSSLFTMDIWRRLRPRSGERELLLVGRV
IVALIGVSVAWIPVLQDSNSGQLFYIMQSVTSSLPVTAFFVLGVFWRANEQGAFFWGLIAGLVVGATRLV
LEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVVAGSLTPTPPQSVQIENLTWWTLAQDVPLGTK
AGDGQTLQKHAFWARVCGFNAILLMCVNIFFYAYFA

A search of sequence databases reveals that the NOV27b amino acid sequence has 530 of 612 amino acid residues (86%) identical to, and 558 of 612 amino acid residues (91%) similar to, the 597 amino acid residue ptnr:SPTREMBL-ACC:Q28610 protein from *Oryctolagus cuniculus* (Rabbit) (Na⁺/Glucose Cotransporter-Related Protein) ($E = 1.9e^{-284}$).

NOV27b is predicted to be expressed in at least heart and kidney. The sequence is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:OCU08813|acc:U08813.1) a closely related *Oryctolagus cuniculus* Na⁺/glucose cotransporter-related protein mRNA, complete cds homolog.

NOV27c

A disclosed NOV27c nucleic acid of 1741 nucleotides (also referred to as 191828203) encoding a novel sodium - glucose cotransporter-like protein is shown in Table 27E. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 5-7 and ending with a TGA codon at nucleotides 1688-1690. The start and stop codons are shown in bold in Table 27E, and the 5' and 3' untranslated regions, if any, are underlined.

Table 27E. NOV27c nucleotide sequence (SEQ ID NO:117).

AGCCATGGCCGCAACTCCACCAGCGACCTCCACACTCCCGGACGCAGCTGAGCGTGGCTGACATCATCGT
CATCACTGTGTATTTGCTCTGAATGTGGCCGTGGGCATATGGTCTCTTGTGCGGGCCAGTAGGAACACGGT
GAATGGCTACTTCTGCGCAGCCCGGACATGACGTGGTGGCCGATGGAGCCTCCCTCTCGCCAGCAGCGA
GGGCTCTGGCCTCTTCATTGGACTGGCGGGCTCAGGCGGGCAGGAGGTCTGGCCGTGGCAGGCTTCGAGTG
GAATGCCACGTACGTGCTGCTGGCACTGGCATGGGTGTTCTGTGCCCATCTACATCTCCTCAGAGCTGGACCT
GTACGCGGGGGCTCTGTTTGTGCACATCTGCCTGGGCTGGAACCTCTACCTCTCCACCATCCTCAGCTCGG
CATCACAGCCCTGTACACCATCGCAGGGGGCTGGCTGCTGTAATCTACACGACGCCCTGCAGACGCTCAT
CATGGTGGTGGGGGCTGTATCCTGACAATCAAAGCTTTTGACCAGATCGGTGGTTACGGGCAGCTGGAGGC
AGCCTACGCCAGGCCATTCCCTCCAGGACCATTGCCAACACCACCTGCCACCTGCCACGTACAGACGCCAT
GCACATGTTTCGAGACCCACACAGGGGACCTGCCGTGGACCGGATGACCTTTGGCCTGACCATCATGGC
CACCTGGTACTGGTGACCGACCGAGGTCATCGTGACGACGATCACTGTCAGCCCGGGACCTGAACCATGCCAA

GGCGGGCTCCATCCTGGCCAGCTACCTCAAGATGCTCCCATGGGCTGATCATCATGCCGGGCATGATCAG
 CCGCGCATTTGTTCCAGATGATGTGGGCTGCGTGGTGCCGTCCGAGTGCCTGCGGGGCTGCGGGGCGGAGGT
 CGGCTGCTCCAACATCGCCTACCCCAAGCTGGTCATGGAATGATGCCATCGGTCTGCGGGGGCTGATGAT
 CACAGTGATGCTGGCGGCGCTCATGTGCTCGCTGACCTCCATCTTCAACAGCAGCAGCACCCCTCTTCACTAT
 GGACATCTGGAGGCGGCTGCGTCCCCGCTCCGGCGAGCGGGAGCTCTGCTGGTGGGACGGCTGGTCATAGT
 GGCACATCATCGGCGTGAGTGTGGCCTGGATCCCCGCTCTGCAGGGCTCCAACAGCGGGCAACTCTTCACTTA
 CATGCAGTCAGTGACCAGCTCCCTGGCCCCACAGTGACTGCAGTCTTTGTCTGGGCGTCTTCCGGCGACG
 TGCCAACGAGCAGGGGGCCTTCTGGGGCTGATAGCAGGGCTGGTGGTGGGGGCCACGAGGCTGGTCTCTGGA
 ATTCCTGAACCCAGCCCCACCGTGCGGAGAGCCAGACACGCGGCCAGCCGTCTGGGGAGCATCCACTACCT
 GCACTTCGCTGTGCGCCTCTTTCAGTCACTCAGTGGTGTCTGTTGTGGTGGCTGGAAGCCTGCTGACCCACCCCC
 ACAGAGTGTCAGATTGAGAACCTTACCTGGTGGACCCCTGGCTCAGGATGTGCCCTTGGGAACATAAGCAGG
 TGATGGCCAAACACCCAGAAACACGCCTTCTGGGCGCGCTCTGTGGCTTCAATGCCATCCTCTCATGTG
 TGTCACATATTCTTTATGCCTACTTCGCCTGACACTGCCATCCTGGACAGAAAGGCAGGAGCTCTGAGTT
 GGCGGCCATGGCT

In a search of public sequence databases, the NOV27c nucleic acid sequence, located on chromosome 17, has 1409 of 1445 bases (97%) identical to a gb:GENBANK-
 ID:AX191622|acc:AX191622.1 mRNA from *Homo sapiens* (Sequence 144 from Patent
 5 WO0149728) (E = 0.0).

A disclosed NOV27c polypeptide (SEQ ID NO:118) encoded by SEQ ID NO:117 has 561 amino acid residues and is presented in Table 27F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV27c has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8200. Alternatively,
 10 NOV27c may also localize to the endoplasmic reticulum (membrane) with a certainty of 0.6850, to the Golgi body with a certainty of 0.4600, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV27C is between positions 42 and 43: CRA-SR.

Table 27F. Encoded NOV27c protein sequence (SEQ ID NO:118).

MAANSTSDLHTPGTQLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGYFLAGRDMTWNPIGASLFASSEG
 SGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVPIYISSELDLYAGALFVHICLGNWFYLSLTILTLGI
 TALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGLEAAYAQAIPSRITANTTCHLPRTDAMH
 MFRDPHTGDLPTWGTMTFGLTIMATWYCTDQVIVQRSLSDRLNHAKAGSILASYLKMLPMGLIIMPGMISR
 ALFPDDVGCVPVSECLRACGAEVGCSNIAYPKLVMEMLPIGLRGLMITVMLAALMSSLTIFNSSSTLFTMD
 IWRRLRPRSGERELLLVGRVLIVALIGVSAWIPVLQGSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFRRRA
 NEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHFAVALFALSGAVVVGSLTTPPPQ
 SVQIENLTWWTLAQDVPLGTAKAGDQTPQKHAFWARVCGFNAILLMCVNIFFYAYFA

A search of sequence databases reveals that the NOV27c amino acid sequence has 394 of 460 amino acid residues (85%) identical to, and 423 of 460 amino acid residues (91%) similar to, the 597 amino acid residue ptrn:SPTREMBL-ACC:Q28610 protein from *Oryctolagus cuniculus* (Rabbit) (Na⁺/Glucose Cotransporter-Related Protein) (E = 2.6e⁻¹²⁵).

NOV27c is predicted to be expressed in at least heart, kidney, and colon. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of CuraGen Acc. No. 191828203. The sequence is predicted to be

expressed in kidney because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:OCU08813|acc:U08813.1) a closely related *Oryctolagus cuniculus* Na⁺/glucose cotransporter-related protein mRNA, complete cds homolog.

The disclosed NOV27a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 27G.

Table 27G. BLAST results for NOV27a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 520469 gb AAA66065.1 (U08813)	597 aa protein related to Na/glucose cotransporters [Oryctolagus cuniculus]	597	531/596 (89%)	559/596 (93%)	0.0
gi 16553933 dbj BAB71619.1 (AK057946)	unnamed protein product [Homo sapiens]	517	440/456 (96%)	440/456 (96%)	0.0
gi 18203958 gb AAH21357.1 AAH21357 (BC021357)	Unknown (protein for MGC:29197) [Mus musculus]	678	346/545 (63%)	435/545 (79%)	0.0
gi 9588428 emb CAC00574.1 (AL109659)	dJ1024N4.1 (novel Sodium:solute symporter family member similar to SLC5A1 (SGLT1)) [Homo sapiens]	552	344/522 (65%)	425/522 (80%)	e-180
gi 2564063 dbj BAA22950.1 (AB008225)	Na ⁺ -glucose cotransporter type 1 (SGLT-1)-like protein [Xenopus laevis]	673	315/539 (58%)	415/539 (76%)	e-174

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 27H. In the ClustalW alignment of the NOV27 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 27H. ClustalW Analysis of NOV27

- 1) Novel NOV27a (SEQ ID NO:114)
- 2) Novel NOV27b (SEQ ID NO:116)
- 3) Novel NOV27b (SEQ ID NO:118)
- 4) gi|520469|gb|AAA66065.1| (U08813) 597 aa protein related to Na/glucose cotransporters [Oryctolagus cuniculus] (SEQ ID NO:436)
- 5) gi|16553933|dbj|BAB71619.1| (AK057946) unnamed protein product [Homo sapiens] (SEQ ID NO:437)

- 6) gi|18203958|gb|AAH21357.1|AAH21357 (BC021357) Unknown (protein for MGC:29197) [Mus musculus] (SEQ ID NO:438)
 7) gi|9588428|emb|CAC00574.1| (AL109659) dJ1024N4.1 (novel Sodium:solute symporter family member similar to SLC5A1 (SGLT1)) [Homo sapiens] (SEQ ID NO:439)
 5 8) gi|2564063|dbj|BAA22950.1| (AB008225) Na⁺-glucose cotransporter type 1 (SGLT-1)-like protein [Xenopus laevis] (SEQ ID NO:440)

			10	20	30	40	50	60	
10	NOV27a	1	-MAANSTSDLHTPGT-----QLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGY	50					
	NOV27b	1	-MAANSTSDLHTPGT-----QLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGY	50					
	NOV27c	1	-MAANSTSDLHTPGT-----QLSVADIIVITVYFALNVAVGIWSSCRASRNTVNGY	50					
	gi 520469	1	MVADNSTSDPHAPGP-----QLSVTDIVVITVYFALNVAVGIWSSCRASRNTVSGY	51					
	gi 16553933	1	-MPRTCCWHWHG-----CSCPS--TS--	18					
15	gi 18203958	1	MEPGVSRNGVRTETT---TNPSLGLHTYDIIVVVVIYFVFLAVGIWSSIRASRGTVGGY	56					
	gi 9588428	1	MGPASGDGVRTETAPHIALDSRVGLHAYDISVVVVIYFVFLAVGIWSSIRASRGTVGGY	60					
	gi 2564063	1	METSSQSSPQTTPGMEAFPKK---SLDTDIIVVVVIYFVFLAVGLLSMCRTRKRGTVKGY	57					
			70	80	90	100	110	120	
20	NOV27a	51	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	110					
	NOV27b	51	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	110					
	NOV27c	51	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	110					
	gi 520469	52	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	111					
25	gi 16553933	18	-----PQRSSPYLSTFRS-----ATGASG-----SACTCLSCPCYCL-	51					
	gi 18203958	57	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	116					
	gi 9588428	61	FLAGRDMTWWPIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	120					
	gi 2564063	58	FLAGKDMANWVPVIGASLSESGSLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFVP	117					
			130	140	150	160	170	180	
30	NOV27a	111	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	170					
	NOV27b	111	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	170					
	NOV27c	111	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	170					
35	gi 520469	112	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	171					
	gi 16553933	51	-----SPRYRWICTR-----GALFVHICLGWNFY	75					
	gi 18203958	117	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	176					
	gi 9588428	121	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	180					
	gi 2564063	118	IYISSEIVTPEYIQRKRGQRIQVMSVLSLIIYVFTKISLDIYAGALFVHICLGWNFY	177					
40			190	200	210	220	230	240	
45	NOV27a	171	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	230					
	NOV27b	171	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	230					
	NOV27c	136	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	195					
	gi 520469	172	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	231					
	gi 16553933	76	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	135					
	gi 18203958	177	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	236					
	gi 9588428	181	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	240					
50	gi 2564063	178	LSTILTITGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIKAFDQIGGYGQLEAAYAQ	237					
			250	260	270	280	290	300	
55	NOV27a	231	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	290					
	NOV27b	231	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	290					
	NOV27c	196	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	255					
	gi 520469	232	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	291					
	gi 16553933	136	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	195					
	gi 18203958	237	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	296					
	gi 9588428	241	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	300					
60	gi 2564063	238	AIPSRTHANTTCHLPRDAMHMFDPHTGDLPTWGTMTFGLTIMATWYWCDDQVIVQRSLS	297					
			310	320	330	340	350	360	
65	NOV27a	291	ARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFP-----DDVGCVV	334					
	NOV27b	291	ARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFP-----DDVGCVV	350					
	NOV27c	256	ARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFP-----DDVGCVV	299					
	gi 520469	292	ARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFP-----DDVGCVV	335					
	gi 16553933	196	ARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFP-----DDVGCVV	255					

gi|18203958| 297 AKNLSHAKGGSVLGGYLKILPMFFIVMPGMISRALMP-----DEVACVD 340
gi|9588428| 301 AKSLSHAKGGSVLGGYLKILPMFFIVMPGMISRALFP-----DEVGCVD 344
gi|2564063| 298 AKNLSHAKAGSLAASLKVLLEIFMMVLPGMISRVLFT-----DOVACAD 341

5 370 380 390 400 410 420

NOV27a 335 PSECLRACGAEVGCSNIAYPKLVMEMLPTGLRGLMIAVMLAALMSSLTSIFNSSSTLFTM 394
NOV27b 351 PSECLRACGAEVGCSNIAYPKLVMEMLPTGLRGLMIAVMLAALMSSLTSIFNSSSTLFTM 410
NOV27c 300 PSECLRACGAEVGCSNIAYPKLVMEMLPTGLRGLMTIVMLAALMSSLTSIFNSSSTLFTM 359
10 gi|520469| 336 PSECLRACGAELIGCSNIAYPKLVMEMLPVGLRGLMIAVMPALMSSLSSIFNSSSTLFTM 395
gi|16553933| 256 PSECLRACGAEVGCSNIAYPKLVMEMLPTGLRGLMIAVMLAALMSSLTSIFNSSSTLFTM 315
gi|18203958| 341 PDICORVCGARVGCNIAYPKLVMAALMPVGLRGLMIAVMAALMSSLTSIFNSSSTLFTM 400
gi|9588428| 345 PDVQORTCGARVGCNIAYPKLVMAALMPVGLRGLMIAVMAALMSSLTSIFNSSSTLFTL 404
15 gi|2564063| 342 PELCKEICGNPSGCSDIAYPKVHIELPTGLRGLMISVMIAALMSSLTSIFNSASTLFTM 401

430 440 450 460 470 480

NOV27a 395 DIWRRRLRPRSGERELLVGRVLVIVALIGSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPV 454
NOV27b 411 DIWRRRLRPRSGERELLVGRVLVIVALIGSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPV 470
NOV27c 360 DIWRRRLRPRSGERELLVGRVLVIVALIGSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPV 419
20 gi|520469| 396 DIWRRRLRPRSGERELLVGRVLVIVALIGSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPV 455
gi|16553933| 316 DIWRRRLRPRSGERELLVGRVLVIVALIGSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPV 375
gi|18203958| 401 DWQQRFRROASEQELMVVGRVFFVFLVVISLWIPIIICSSNSGQLFDYIQSITSYLAPPF 460
gi|9588428| 405 DWQQRFRRKSTEQELMVVGRVFFVFLVVISLWIPIIICSSNSGQLFDYIQAVTSYLAPPF 464
25 gi|2564063| 402 DLWRHTRPRSTEWELMVGRVFFVFLVVISLWIPIIICASOGGQLFIYIQSESSYLQPPV 461

490 500 510 520 530 540

NOV27a 455 TAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYL 514
NOV27b 471 TAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYL 530
NOV27c 420 TAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYL 479
30 gi|520469| 456 TAVFTLGIFFWRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYL 515
gi|16553933| 376 TAVFVLGVFWRRANEQGAFWGLIAGLVVGATRLVLEFLNPAPPCGEPDTRPAVLGSIHYL 435
gi|18203958| 461 TAFEFLLAIFCKRVNEPGAFAWGLMFLVVGILRMILEFSYAPACGEMDRRPAVLKDFHYL 520
35 gi|9588428| 465 TAFEFLLAIFCKRVTEPGAFAWGLVFGICVGLLRMILEFSYPAPACGEVDRRPAVLKDFHYL 524
gi|2564063| 462 AVVFAGCFWKRINEKGAFWGMTIIGLVVGLIRMVLDFTIYVAPQCDQPDTRPGVVKYVHYL 521

550 560 570 580 590 600

NOV27a 515 HFAVALFALS GAVVVAGSLLTPPPQSVQIENLTWWT----- 550
NOV27b 531 HFAVALFALS GAVVVAGSLLTPPPQSVQIENLTWWT----- 566
NOV27c 480 HFAVALFALS GAVVVAGSLLTPPPQSVQIENLTWWT----- 515
40 gi|520469| 516 HFAVALEVLITGAVAVGSSLLTPPPQSVQIENLTWWT----- 551
gi|16553933| 436 HFAVALFALS GAVVVAGSLLTPPPQSVQIENLTWWT----- 471
45 gi|18203958| 521 YFALLCGLTAITIVVISFTEPIPDCKIARLTWWTTRNCAVSDLQKKT-----SVSVN 573
gi|9588428| 525 YFALLCGLTAITIVVISLCTTPIPEEC----- 552
gi|2564063| 522 YLSMITGLLTLLVVVAVSLWTEPPSKQMISRLTWETRFDAEVEDPVTNHRPAENGISVV 581

610 620 630 640 650 660

NOV27a 550 -----LAQDVPLGTKAGDGQTP----- 567
NOV27b 566 -----LAQDVPLGTKAGDGQTL----- 583
NOV27c 515 -----LAQDVPLGTKAGDGQTP----- 532
50 gi|520469| 551 -----LTRDLSLGAAGDGQTP----- 568
55 gi|16553933| 471 -----LAQDVPLGTKAGDGQTP----- 488
gi|18203958| 574 NTEGDNSPGLAGRPVVEGPGAGDEEEANTTQGEQPGALHRSWGKWLWNWFCGLSGAPQQA 633
gi|9588428| 552 ----- 552
gi|2564063| 582 EDISEEPHTTSTDAIYNDTTDNPSS-S-----SLLKKTILWLCGMDSRKGDG 628

670 680 690 700

NOV27a 567 -----OKHAFWARVCGFNAILLMCVNIFFYAYFA 596
NOV27b 583 -----OKHAFWARVCGFNAILLMCVNIFFYAYFA 612
NOV27c 532 -----OKHAFWARVCGFNAILLMCVNIFFYAYFA 561
60 gi|520469| 568 -----QRYTFWARVCGFNAILLMCVNIFFYAYFA 597
gi|16553933| 488 -----OKHAFWARVCGFNAILLMCVNIFFYAYFA 517
gi|18203958| 634 LSPAFAKAVLEQKLTSTIEEPLWRRVONTNAILLLANINIFLWGYFS 678
gi|9588428| 552 ----- 552
65 gi|2564063| 629 HDQAPPAPLEPAEVLVYERPLLKQVLTAVTLCMSAGVELWAYFG 673

Table 27I lists the domain description from DOMAIN analysis results against NOV27. This indicates that the NOV27 sequence has properties similar to those of other proteins known to contain this domain.

Table 27I Domain Analysis of NOV27

gnl|Pfam|pfam00474, SSF, Sodium:solute symporter family. (SEQ ID NO:820)
CD-Length = 406 residues, 100.0% aligned
Score = 310 bits (793), Expect = 2e-85

NOV27:	50	YFLAGRDMTWWPIGASLFASSEGSGLFIGLAGSGAAGGLAVAGFEWNATYVLLALAWVFV	109
Sbjct:	1	YFLAGRSMTGFVNGLSLAASYMSAASFVGLAGAGAASGLAGGLYAIGALVGVWLLWLFA	60
NOV27:	110	PIYISSEIVTLPEYIQKRYGGQIRMYLSVLSLLSVFTKISLDLYAGALFVHICLGWNF	169
Sbjct:	61	PRLRNLGAYTMPDYLRKRFGGKRILVYLSALSLLYFFTYMSVQIVGGARLIELALGLNY	120
NOV27:	170	YLSTILTLGITALYTIAGGLAAVIYTDALQTLIMVVGAVILTIFAFDQIGGYGQLEAAYA	229
Sbjct:	121	YTAVLLLGALTAIYTFGGFLAVSWTDTIQAVLMLFGTIILMIIVFHEVGGYSSAVEKYM	180
NOV27:	230	QAIPSRITANTTCHLPRTDAMHMRDPHTGDLPTWGTMTFGLTIMATWYWCCTDQVIVQRSL	289
Sbjct:	181	TADPNGVDLYT-----PDGLHILRDPLTGLSLWPLGLVLGTTGL-----PHILQRCL	226
NOV27:	290	SARDLNHAKAGSILASYLKMLPMGLIIMPGMISRALFPDDVGCVPSECLRACGAEVGCS	349
Sbjct:	227	AAKD-----AKCIRCGVLILTPMFIIVMPGMISRGFLAIALAGANP----RACGTVVGCS	277
NOV27:	350	NIAYPKLVMELMPIGLRGLMIAVMLAALMSSLTSIFNSSSTLFTMDIWRRLRPRSGEREL	409
Sbjct:	278	NIAYPTLAVKLGPPGLAGIMLAAMSTLTSQLLSSSSAFTHDLYKNIRRKASATEK	337
NOV27:	410	LLVGRLVIVALIGVSVAVIPVLQDSNSGQLFIYMQSVTSSLAPPVTAVFVLGVFWRRANE	469
Sbjct:	338	ELVGRSRIIVLVVISLAILLAVQ-PAQMGIAPLVQLAFAGLGSAPLPVILLAIFWKRVNE	396
NOV27:	470	QGAFWGLIAG	479
Sbjct:	397	QGALWGMIIG	406

The gene of invention codes for a human ortholog of a rabbit sodium-glucose cotransporter (SGLT) and belongs to the large family of SGLTs that has been described to date. The rabbit gene is expressed in the kidney (Pajor, Biochim Biophys Acta 1994 Sep 14;1194(2):349-51), and the novel gene described herein is expressed in the heart in addition to the kidney. It shows the characteristic sodium-solute symporter protein motif shared by members of the SGLT family.

SGLTs are critical in the maintenance of glucose homeostasis in the body, in a variety of tissues. Inhibitors of SGLTs are being studied in the treatment of diabetes. Treatment of

Zucker diabetic fatty rats with the SGLT inhibitor T-1095 lowers both fed and fasted blood glucose levels to near-normal levels (Nawano et al., Am J Physiol Endocrinol Metab 2000 Mar;278(3):E535-43). In streptozotocin-induced diabetic rats, T-1095 also exerts an antihyperglycemic effect which is nullified by nephrectomy, indicating that the drug acts
5 through inhibition of renal SGLTs rather than intestinal ones (Oku et al., Biol Pharm Bull 2000 Dec;23(12):1434-7) In addition, SGLT-1 seems to have a role in mammalian renal tubulogenesis (Yang et al., Am J Physiol Renal Physiol 2000 Oct;279(4):F765-77).

The disclosed NOV27 nucleic acid of the invention encoding a Sodium - Glucose Cotransporter -like protein includes the nucleic acid whose sequence is provided in Table 27A,
10 27C, 27E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 27A, 27C, or 27E while still encoding a protein that maintains its Sodium - Glucose Cotransporter -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described,
15 including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
20 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 14 percent of the bases may be so changed.

The disclosed NOV27 protein of the invention includes the Sodium - Glucose
25 Cotransporter -like protein whose sequence is provided in Table 27B, 30D, or 30F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 27B, 27D, or 27F while still encoding a protein that maintains its Sodium - Glucose Cotransporter -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 42
30 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Sodium - Glucose Cotransporter - like protein (NOV27) is a member of a "Sodium - Glucose Cotransporter family". Therefore,

the NOV27 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug
5 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV27 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in diabetes, obesity, hypertension, cardiomyopathy,
10 atherosclerosis, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, transplantation, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular
15 acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, cancer, tissue degeneration, diabetic nephropathy, microvascular and macrovascular disease, and/or other diseases and pathologies.

NOV27 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV27 substances for use in
20 therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV27 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in
25 understanding of pathology of the disease and development of new drug targets for various disorders.

NOV28

A disclosed NOV28 nucleic acid of 1560 nucleotides (also referred to as CG56185-01) encoding a MYD-I-like protein is shown in Table 28A. An open reading frame was identified
30 beginning with a ATG initiation codon at nucleotides 31-33 and ending with a TGA codon at nucleotides 1537-1539. The start and stop codons are shown in bold in Table 28A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 28A. NOV28 nucleotide sequence (SEQ ID NO:119).

CAGCCCTCGCGGGCGCGTAGCCGCGGCCCATGGAGCCCGCGGGCCGGTCCCCGGCCGCTCGGGCCGCTGC
TCTGCCTGCTGCTCCCCGCGTCTGCGCCTGGTGGAGTGGCGGGTGAGGAGGAGCTGCAGGTGATTCAGCCT
GAGAAGTCTGTATCAGTTGAGCTGGAGAGTTCGGCCGCTCTGCAGTGCAGTGTGACCTCCCTGAACCTGTG
GGGCCCATCCAACGGTTCAGAGGAGCTGGACCAGGCCGGAATTAATCTACCATCAAAAAGAAGGCCACTTC
CCCCGGGTAACTGTTTTCAGATCTCACAAAGAGAACCAACATGGACTTTTCCATCTGCATCAGTAACATC
ACCCAGCAGATGCCGGCACCTACTACTGTGTGAAGTTCCAGAAAGGAGCCCTGACGTGGAGTTGAAGTCT
GGAGCAGGCACTGAGCTGTCTGTGCGTGCCAAACCTCTGCCCCCGTGGTATCGGGCCCCGAGCGAGGGCC
ACACCTGACCACACAGTGTGAGCTTACCTGCGAGTCTCATGGCTTCTCACCCAGAGACATCAGCCTGAAATGG
TTCAAAAATGGGAATCAGCTCTCAGACTTCCAGACCAACGTGGACCCCGAAGAGAGAGCGTGTCTACAGC
ATCCACAGCACAGCCAATGTGGTGTGACCCGCGGGGACATTCACTCTCAAGTCATCTGCCAGGTGGCCAC
GTCACCTTGCGGGGGGACTCTTTTCTGTTGGGACTGCCAATTTGTCTGAGACTATCCAAGTTCACCCACCTTG
GAGGTTACTCAACAGCCATGAGGGCAGAGAACAGGTGAATATCACTGCCAGGTGACGAAATTTACCC
CAGAGACTACAGTTGACCTGGTTGGAGAACGCAATGTGTCCCGACAGAAACGGCCTCAACTCTTACAGAG
AACAGGATGGCACCTTACAACCTGGATGAGTGGCTCCTGGTGAATGTATCTGCCACAGGGATGATGTGAAG
CTCACCTGCCAGGTGGAGCATGACGGGCAGTCAGCGGTGACGAAAAGCCATGACCTGAAGGTCTCAGCCAC
CTGAAGGAGCAGAGCTCAAATACCGCCGCTGAGAACACTGGACCTAATGAACAGAACATCTATATTGGTGTG
GGCGTGGTGTGCACCTTGTGTGGTGGCCCTACTGATGGAGGCTCTCTACCTCGTCCGAATCAGACAGAGAAA
GCCAGGGCTCACTTCTTCTACAAGTTGATGAACCCGAGAAGATGCCAGAAAAATAACCCAGGACACA
AATGATATCACATATGCGGACCTGAACCTGCCAAGGGGAAGAAGCTGCTCCCCGGGGCCGCGAGCCCAAC
AACCACACAGAGTATGCCAGCATTAGACCAGCCTGCAGCCTGCGTCCGAGGACACCCCTACCTATGCTGAC
CTGGACATGGTGCACCTCAACCGGACCCCAAGCAGCTGGCCCCCAAGCCCGAGCTGTCCTTCTCAGAGTAT
GCCAGCATCCAGGTCCCGAGGAAGTGAATGGGACCGTGGTTTGTCTCTA

In a search of public sequence databases, the NOV28 nucleic acid sequence, located on chromosome 22, has 1466 of 1544 bases (94%) identical to a gb:GENBANK-
ID:HSSIRPALP|acc:Y10375.1 mRNA from *Homo sapiens* (*H.sapiens* mRNA for SIRP-
5 alpha1) ($E = 7.4e^{-310}$).

The disclosed NOV28 polypeptide (SEQ ID NO:120) encoded by SEQ ID NO:119 has 503 amino acid residues and is presented in Table 28B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV28 has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.4600. Alternatively,
10 NOV28 may also localize to the endoplasmic reticulum (membrane) with a certainty of 0.1000, to the endoplasmic reticulum (lumen) with a certainty of 0.1000, or extracellularly with a certainty of 0.1000. The most likely cleavage site for NOV28 is between positions 30 and 31: VAG-EE.

Table 28B. Encoded NOV28 protein sequence (SEQ ID NO:120).

MEPAGVPVGRLLGPLLCLLLPASCAWSGVAGEEELQVIQPEKSVSVAAGESAALQCTVTSLNVPVGPIQRFRTGA
GPGRKLIYHQKEGHFPRVTTVSDLTTRKTNMDFISICISNITPADAGTYICVKFKGSPDVELKSGAGTELSVR
AKPSAPVVS GPAARATPDHTVSFTCESHGFSRPDISLKWFKNGNQLSDFQTNVDPARESVSYSIHSTANVVL
TRGDIHSQVICEVAHVTLRGDSFRGTANLSETIQVPPLEVTQQPMRAENQVNITCQVTKFYQPORLQLTWLE
NGNVSRTEASTLTENKDGTYNWSWLLVNVAHRDDVKLTQVEHDGQSASVSKSHDLKVS AHLKEQSSNTA
AENTGPNEQNIYIVGVVCTLLVALLMEALYLVRIKKAQGSTSSTRLHEPEKNARKITQDNDITYADLN
LPKGKKPAPRAAEPNNHTEYASIQTSLQPASEDTLTYADLDMVHLNRTPKQLAPKPELSFSEYASIQVPRK

15

A search of sequence databases reveals that the NOV28 amino acid sequence has 458 of 503 amino acid residues (91%) identical to, and 475 of 503 amino acid residues (94%) similar to, the 503 amino acid residue ptrn:SPTREMBL-ACC:P78324 protein from *Homo*

sapiens (Human) (Protein Tyrosine Phosphatase, Non-Receptor Type Substrate 1 Precursor (Shp Substrate-1) (Inhibitory Receptor Shps-1) (Shps-1) (Signal- Regulatory Protein Alpha-1) (SIRP-Alpha1) (MYD-1 Antigen)) ($E = 5.7e^{-247}$).

NOV28 is predicted to be expressed in at least myeloid, macrophages, Adrenal Gland/Suprarenal gland, Bone Marrow, Brain, Whole Organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in myeloid and macrophages because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSSIRPALP| acc: Y10375.1) a closely related *H.sapiens* mRNA for SIRP-alpha1 homolog.

The disclosed NOV28 polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 28C.

Table 28C. BLAST results for NOV28					
Gene Index/ Identifier	Protein/ Organism	Length (aa).	Identity (%)	Positives (%)	Expect
gi 14771369 ref XP_044897.1 (XM_044897)	hypothetical protein XP_044897 [<i>Homo sapiens</i>]	504	458/504 (90%)	476/504 (93%)	0.0
gi 4758978 ref NP_04639.1 (NM_004648)	protein tyrosine phosphatase, non- receptor type substrate 1; SHP substrate-1 [<i>Homo sapiens</i>]	503	458/503 (91%)	475/503 (94%)	0.0
gi 6624134 gb AAF19260.1 AC004832_5 (AC004832)	similar to SHPS-1 [<i>Homo sapiens</i>]; similar to BAA12974.1 (PID:g1864011)	402	402/402 (100%)	402/402 (100%)	0.0
gi 2842392 emb CAA71944.1 (Y11047)	MyD-1 antigen [<i>Homo sapiens</i>]	429	391/429 (91%)	407/429 (94%)	0.0
gi 2842390 emb CAA71942.1 (Y11045)	MyD-1 antigen [<i>Bos taurus</i>]	506	373/510 (73%)	415/510 (81%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 28D. In the ClustalW alignment of the NOV28 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 28D. ClustalW Analysis of NOV28

1) Novel NOV28 (SEQ ID NO:120)	
5	2) gi 14771369 ref XP_044897.1 (XM_044897) hypothetical protein XP_044897 [<i>Homo sapiens</i>] (SEQ ID NO:441)
	3) gi 4758978 ref NP_004639.1 (NM_004648) protein tyrosine phosphatase, non-receptor type substrate 1; SHP substrate-1 [<i>Homo sapiens</i>] (SEQ ID NO:442)
10	4) gi 6624134 gb AAF19260.1 AC004832_5 (AC004832) (SEQ ID NO:443)
	5) similar to SHPS-1 [<i>Homo sapiens</i>]; similar to BAA12974.1 (PID:g1864011) (SEQ ID NO:444)
15	6) gi 2842392 emb CAA71944.1 (Y11047) MyD-1 antigen [<i>Homo sapiens</i>] (SEQ ID NO:445)
	7) gi 2842390 emb CAA71942.1 (Y11045) MyD-1 antigen [<i>Bos taurus</i>] (SEQ ID NO:446)
20	NOV28 1 MEPAGFPVPGRLGPLLCLLLPASCWSGVAGEEELQVIOPEKSVSVAAGESAALQCTVTSL 60
	gi 14771369 1 MEPAGFPVPGRLGPLLCLLLAASCWSGVAGEEELQVIOPEKSVLVAAGETATLRCTATSL 60
25	gi 4758978 1 MEPAGFPVPGRLGPLLCLLLAASCWSGVAGEEELQVIOPEKSVSVAAGESAALHCTVTSL 60
	gi 6624134 1 ----- 1
30	gi 2842392 1 ----- 1
	gi 2842390 1 MEPAFPVPGRLGPLLCLLLAASSAWTCAAGDELOVIOPEKSVSVAAGETATLRHCTVTSL 60
35	NOV28 61 NPVGPIQRFRGAGPGRKLIYHQKEGHFPRVTTVSDLTKRNNMDFSICISNITPADAGTYV 120
	gi 14771369 61 IPVGPIQWFRGAGPGRRELIYVNOKEGHFPRVTTVSDLTKRNNMDFSIRIGNITPADAGTYV 120
40	gi 4758978 61 IPVGPIQWFRGAGPARELIYVNOKEGHFPRVTTVSESTKRNNMDFSISISNITPADAGTYV 120
	gi 6624134 1 -----MDFSICISNITPADAGTYV 19
45	gi 2842392 1 -----RELIVNNOKEGHFPRVTTVSDLTKRNNMDFSIRIGNITPADAGTYV 45
	gi 2842390 61 LPVGPIKWFRTGPGREFIYVNOKEAFPRVTNVSDATKRNNMDFSIRISNITPADAGTYV 120
50	NOV28 121 CVKFKKGSPDVELKSGAGTELSVRAKPSAPVVS GPAARATPDHTVSFTCESHGFSRDI 179
	gi 14771369 121 CVKFKKGSPDDVEFKSGAGTELSVRAKPSAPVVS GPAARATPDHTVSFTCESHGFSRDI 180
55	gi 4758978 121 CVKFKKGSPDTEFKSGAGTELSVRAKPSAPVVS GPAARATPDHTVSFTCESHGFSRDI 179
	gi 6624134 20 CVKFKKGSPDVELKSGAGTELSVRAKPSAPVVS GPAARATPDHTVSFTCESHGFSRDI 78
60	gi 2842392 46 CVKFKKGSPDDVEFKSGAGTELSVRAKPSAPVVS GPAARATPDHTVSFTCESHGFSRDI 105
	gi 2842390 121 CVKFKKGEGHGDVEFKSGPGTGLIVNAKPSPPVVS GPATVRATEPQTVNFTCTSHGFSRDI 180
65	NOV28 180 SLKWFKNGNQLSDFQTNVDPARESVSYSIHSTANVVLTRGDTHSQVICEVAHVTLRGDS 238
	gi 14771369 181 TLKWFKNGNELSDFTQNVDPGCVESVSYSIHSTAKVVLTRGDVHSQVICEVAHVTLQGDP 239
70	gi 4758978 180 TLKWFKNGNELSDFTQNVDPGCVESVSYSIHSTAKVVLTRGDVHSQVICEVAHVTLQGDP 238
	gi 6624134 79 SLKWFKNGNQLSDFQTNVDPARESVSYSIHSTANVVLTRGDTHSQVICEVAHVTLRGDS 137
75	gi 2842392 106 TLKWFKNGNELSDFTQNVDPGCVESVSYSIHSTAKVVLTRGDVHSQVICEVAHVTLQGDP 164
	gi 2842390 181 SLKWFKNGNELSSAQTSVDPEDDNVSYSINSTTKVILLATGDVHSQVICEVAHVTLQGDP 240
80	NOV28 239 FRGTANLSETIQVPPTLEVTOQPMRAENQVNIITCOVTKFYPRQLQLTLWLENGNVSRTEA 298
	gi 14771369 240 LRGTANLSETIRVPPTLEVTOQPMRAENQVNVTCQVRKFYPRQLQLTLWLENGNVSRTEA 299
85	gi 4758978 239 LRGTANLSETIRVPPTLEVTOQPMRAENQVNIITCOVTKFYPRQLQLTLWLENGNVSRTEA 298
	gi 6624134 138 FRGTANLSETIQVPPTLEVTOQPMRAENQVNIITCOVTKFYPRQLQLTLWLENGNVSRTEA 197
90	gi 2842392 165 LRGTANLSETIRVPPTLEVTOQPMRAENQVNVTCQVRKFYPRQLQLTLWLENGNVSRTEA 224
	gi 2842390 241 LRGTANLSETIRVPPTLEITRSPS-AGNQVNVTCQVKNFYPRQLQLTLWLENGNVSRTEA 299
95	NOV28 299 STLTENKDGTYNWMSSWLLVNVSAHRDDVKLTQVEHGDGQSAVSKSHDLKVS AHPKEQSSN 358
	gi 14771369 300 STVTENKDGTYNWMSSWLLVNVSAHRDDVKLTQVEHGDGQPAVSKSHDLKVS AHPKEQSSN 359
100	gi 4758978 299 STVTENKDGTYNWMSSWLLVNVSAHRDDVKLTQVEHGDGQPAVSKSHDLKVS AHPKEQSSN 358
	gi 6624134 198 STLTENKDGTYNWMSSWLLVNVSAHRDDVKLTQVEHGDGQSAVSKSHDLKVS AHPKEQSSN 257
105	gi 2842392 225 STVTENKDGTYNWMSSWLLVNVSAHRDDVKLTQVEHGDGQPAVSKSHDLKVS AHPKEQSSN 284
	gi 2842390 300 SVTLENKDGTYNMTSWLLVNVSSAHREAVLTQVEHGDGQPAVSKSHDLKVS AHPKEQSSN 359

+ + + +

Sbjct: 61 DVYTCQVTHEGL 72

Protein tyrosine phosphatases (PTPases), such as SHP-1 and SHP-2, that contain Src
homology 2 (SH2) domains play important roles in growth factor and cytokine signal
transduction pathways. A protein of approximately 115 to 120 kDa that interacts with SHP-1
and SHP-2 was purified from v-src-transformed rat fibroblasts (SR-3Y1 cells), and the
corresponding cDNA was cloned. The predicted amino acid sequence of the encoded protein,
termed SHPS-1 (SHP substrate 1), suggests that it is a glycosylated receptor-like protein with
three immunoglobulin-like domains in its extracellular region and four YXX(L/V/I) motifs,
potential tyrosine phosphorylation and SH2-domain binding sites, in its cytoplasmic region.
Various mitogens, including serum, insulin, and lysophosphatidic acid, or cell adhesion
induced tyrosine phosphorylation of SHPS-1 and its subsequent association with SHP-2 in
cultured cells. Thus, SHPS-1 may be a direct substrate for both tyrosine kinases, such as the
insulin receptor kinase or Src, and a specific docking protein for SH2-domain-containing
PTPases. In addition, we suggest that SHPS-1 may be a potential substrate for SHP-2 and may
function in both growth factor- and cell adhesion-induced cell signaling. (Fujioka et al. Mol
Cell Biol. 1996 Dec;16(12):6887-99.)

The rat OX41 antigen is a cell surface protein containing three immunoglobulin
superfamily domains and intracellular immunoreceptor tyrosine-based inhibitory motifs
(ITIM). It is a homologue of the human signal-regulatory protein (SIRP) also known as SHPS-
1, BIT or MFR. Cell activation-induced phosphorylation of the intracellular ITIM motifs
induces association with the tyrosine phosphatases SHP-1 and SHP-2. To identify the
physiological OX41 ligand, recombinant OX41-CD4d3+4 fusion protein was coupled to
fluorescent beads to produce a multivalent cell binding reagent. The OX41-CD4d3+4 beads
bound to thymocytes and concanavalin A-stimulated splenocytes. This interaction was blocked
by the monoclonal antibody (mAb) OX101. Affinity chromatography with OX101 mAb and
peptide sequencing revealed the rat SIRP ligand to be CD47 (integrin-associated protein). A
direct interaction between human SIRP and human CD47 was demonstrated using purified
recombinant proteins and surface plasmon resonance ruling out the involvement of other
proteins known to be associated with CD47. The affinity of the SIRP/CD47 interaction was
K(d) approximately 8 microM at 37 degrees C with a k(off) >= 2.1 s(-1). The membrane-distal
SIRP V-like domain was sufficient for binding to CD47. (Vernon-Wilson EF, et al. Eur J
Immunol. 2000 Aug;30(8):2130-7.)

The transmembrane glycoprotein SHPS-1 binds the protein tyrosine phosphatase SHP-2 and serves as its substrate. Although SHPS-1 has been implicated in growth factor- and cell adhesion-induced signaling, its biological role has remained unknown. Fibroblasts homozygous for expression of an SHPS-1 mutant lacking most of the cytoplasmic region of this protein exhibited increased formation of actin stress fibers and focal adhesions. They spread more quickly on fibronectin than did wild-type cells, but they were defective in subsequent polarized extension and migration. The extent of adhesion-induced activation of Rho, but not that of Rac, was also markedly reduced in the mutant cells. Activation of the Ras-extracellular signal-regulated kinase signaling pathway and of c-Jun N-terminal kinases by growth factors was either unaffected or enhanced in the mutant fibroblasts. These results demonstrate that SHPS-1 plays crucial roles in integrin-mediated cytoskeletal reorganization, cell motility and the regulation of Rho, and that it also negatively modulates growth factor-induced activation of mitogen-activated protein kinases. (Inagaki, A. et al., EMBO J. 2000 Dec 15;19(24):6721-31.)

Machida K. et al. (Oncogene. 2000 Mar 23;19(13):1710-8.) investigated the effect of cell transformation by v-src on the expression and tyrosine phosphorylation of SHPS-1, a putative docking protein for SHP-1 and SHP-2. They found that transformation by v-src virtually inhibited the SHPS-1 expression at mRNA level. While nontransforming Src kinases including c-Src, nonmyristoylated forms of v-Src had no inhibitory effect on SHPS-1 expression, transforming Src kinases including wild-type v-Src and chimeric mutant of c-Src bearing v-Src SH3 substantially suppressed the SHPS-1 expression. In cells expressing temperature sensitive mutant of v-Src, suppression of the SHPS-1 expression was temperature-dependent. In contrast, tyrosine phosphorylation of SHPS-1 was rather activated in cells expressing c-Src or nonmyristoylated forms of v-Src. SHPS-1 expression in SR3Y1 was restored by treatment with herbimycin A, a potent inhibitor of tyrosine kinase, or by the expression of dominant negative form of Ras. Contrary, active form of Mek1 markedly suppressed SHPS-1 expression. Finally, overexpression of SHPS-1 in SR3Y1 led to the drastic reduction of anchorage independent growth of the cells. Taken together, their results suggest that the suppression of SHPS-1 expression is a pivotal event for cell transformation by v-src, and the Ras-MAP kinase cascade plays a critical role in the suppression.

SHPS-1 (SH2-domain bearing protein tyrosine phosphatase (SHP) substrate-1), a member of the inhibitory-receptor superfamily that is abundantly expressed in macrophages and neural tissue, appears to regulate intracellular signaling events downstream of receptor protein-tyrosine kinases and integrin-extracellular matrix molecule interactions. To investigate

the function of SHPS-1 in a hematopoietic cell line, SHPS-1 was expressed in Ba/F3 cells, an IL-3-dependent pro-B-cell line that lacks endogenous SHPS-1 protein. Interestingly, expression of either SHPS-1, or a mutant lacking the intracellular domain of SHPS-1 (DeltaCT SHPS-1), resulted in the rapid formation of macroscopic Ba/F3 cell aggregates. As the integrin-associated protein/CD47 was shown to be a SHPS-1 ligand in neural cells, Babic, J. et al. (J Immunol. 2000 Apr 1;164(7):3652-8.) investigated whether CD47 played a role in the aggregation of SHPS-1-expressing Ba/F3 cells. In support of this idea, aggregate formation was inhibited by an anti-CD47 Ab. Furthermore, erythrocytes from control, but not from CD47-deficient mice, were able to form rosettes on SHPS-1-expressing Ba/F3 cells. Because erythrocytes do not express integrins, this result suggested that SHPS-1-CD47 interactions can take place in the absence of a CD47-integrin association. They also present evidence that the amino-terminal Ig domain of SHPS-1 mediates the interaction with CD47. Although SHPS-1-CD47 binding likely triggers bidirectional intracellular signaling processes, these results demonstrate that this interaction can also mediate cell-cell adhesion.

Inhibitory immunoreceptors downregulate signaling by recruiting Src homology 2 (SH2) domain-containing tyrosine and/or lipid phosphatases to activating receptor complexes [1]. There are indications that some inhibitory receptors might also perform other functions [2] [3]. In adherent macrophages, two inhibitory receptors, SHPS-1 and PIR-B, are the major proteins binding to the tyrosine phosphatase SHP-1. SHPS-1 also associates with two tyrosine-phosphorylated proteins (pp55 and pp130) and a protein tyrosine kinase [4]. Here, Timms, JF. et al. (Curr Biol. 1999 Aug 26;9(16):927-30.) have identified pp55 and pp130 as the adaptor molecules SKAP55hom/R (Src-kinase-associated protein of 55 kDa homologue) and FYB/SLAP-130 (Fyn-binding protein/SLP-76-associated protein of 130 kDa), respectively, and the tyrosine kinase activity as PYK2. Two distinct SHPS-1 complexes were formed, one containing SKAP55hom/R and FYB/SLAP-130, and the other containing PYK2. Recruitment of FYB/SLAP-130 to SHPS-1 required SKAP55hom/R, whereas PYK2 associated with SHPS-1 independently. Formation of both complexes was independent of SHP-1 and tyrosine phosphorylation of SHPS-1. Finally, tyrosine phosphorylation of members of the SHPS-1 complexes was regulated by integrin-mediated adhesion. Thus, SHPS-1 provides a scaffold for the assembly of multi-protein complexes that might both transmit adhesion-regulated signals and help terminate such signals through SHP-1-directed dephosphorylation. Other inhibitory immunoreceptors might have similar scaffold-like functions.

SHPS-1 (or SIRP) is a member of the immunoglobulin (Ig) superfamily abundantly expressed in neurons and other cell types. Within its cytoplasmic domain, it possesses at least

two immunoreceptor tyrosine-based inhibitory motifs, which are targets for tyrosine phosphorylation and mediate the recruitment of SHP-2, an Src homology 2 (SH2) domain-containing protein-tyrosine phosphatase. Since other immunoreceptor tyrosine-based inhibitory motifs-containing receptors have critical roles in the negative regulation of hemopoietic cell functions, the expression of SHPS-1 in cells of hematological lineages was examined. By analyzing a panel of hemopoietic cell lines, evidence was provided that SHPS-1 is abundantly expressed in macrophages and, to a lesser extent, in myeloid cells. No expression was detected in T-cell or B-cell lines. Expression of SHPS-1 could also be documented in normal ex vivo peritoneal macrophages. Further studies showed that SHPS-1 was an efficient tyrosine phosphorylation substrate in macrophages. However, unlike in non-hemopoietic cells, tyrosine-phosphorylated SHPS-1 in macrophages associated primarily with SHP-1 and not SHP-2. Finally, analyses allowed identification of several isoforms of SHPS-1 in mouse cells. In part, this heterogeneity was due to differential glycosylation of SHPS-1. Additionally, it was caused by the production of at least two distinct shps-1 transcripts, coding for SHPS-1 polypeptides having different numbers of Ig-like domains in the extracellular region. Taken together, these findings indicate that SHPS-1 is likely to play a significant role in macrophages, at least partially as a consequence of its capacity to recruit SHP-1. Veillette, A. et al. (J Biol Chem. 1998 Aug 28;273(35):22719-28.)

SHPS-1 is a 120 kDa glycosylated receptor-like protein that contains immunoglobulin-like domains in its extracellular region and four potential tyrosine phosphorylation for SH2 domain binding sites in its cytoplasmic region. Epidermal growth factor (EGF) stimulated the rapid tyrosine phosphorylation of SHPS-1 and subsequent association of SHPS-1 with SHP-2, a protein tyrosine phosphatase containing SH2 domains, in Chinese hamster ovary cells overexpressing human EGF receptors. In the cells overexpressing SHPS-1, the tyrosine phosphorylation of SHPS-1 was more evident than that observed in parent cells. However, overexpression of SHPS-1 alone did not affect the activation of MAP kinase in response to EGF. These results suggest that SHPS-1 may be involved in the recruitment of SHP-2 from the cytosol to the plasma membrane in response to EGF. Copyright 1997 Academic Press. Ochi, F. et al. (Biochem Biophys Res Commun. 1997 Oct 20;239(2):483-7.)

The immune system recognizes invaders as foreign because they express determinants that are absent on host cells or because they lack 'markers of self' that are normally present. Oldenborg et al. (2000) demonstrated that CD47 functions as a marker of self on murine red blood cells. Red blood cells that lack CD47 were rapidly cleared from the bloodstream by splenic red pulp macrophages. CD47 on normal red blood cells prevented this elimination by

binding to the inhibitory receptor signal regulatory protein alpha (SIRP-alpha). Thus, Oldenborg et al. (2000) concluded that macrophages may use a number of nonspecific activating receptors and rely on the presence or absence of CD47 to distinguish self from foreign. Oldenborg et al. (2000) suggested that CD47-SIRP-alpha may represent a potential pathway for the control of hemolytic anemia.

The disclosed NOV28 nucleic acid of the invention encoding a MYD-1 -like protein includes the nucleic acid whose sequence is provided in Table 28A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 28A while still encoding a protein that maintains its MYD-1 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 6 percent of the bases may be so changed.

The disclosed NOV28 protein of the invention includes the MYD-1 -like protein whose sequence is provided in Table 28B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 28B while still encoding a protein that maintains its MYD-1 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 27 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this MYD-1 -like protein (NOV28) is a member of a "MYD-1 family". Therefore, the NOV28 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic

and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV28 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, hemolytic anemia, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, and/or other diseases and pathologies.

NOV28 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV28 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV28 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV29

NOV29 includes three novel CRAL-TRIO -like proteins disclosed below. The disclosed sequences have been named NOV29a, NOV29b, and NOV29c.

NOV29a

A disclosed NOV29a nucleic acid of 1327 nucleotides (also referred to as CG56187-01) encoding a CRAL-TRIO-like protein is shown in Table 29A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 16-18 and ending with a TGA codon at nucleotides 1261-1263. The start and stop codons are shown in bold in Table 29A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 29A. NOV29a nucleotide sequence (SEQ ID NO:121).

<p>GGAGTTGACTGGTGGATGATGTGGGAAGGGTTAGGGGCGGGTGGTGGCCCCGAGGTCATGAGAGCTCCG CCGACCATCAGATCCTCCTCCGCTCAGTTCCGGGAGAACCTCCAGGACCTGCTGCCCATACTGCCCAATGCT GATGACTACTTCCTCCTGCGCTGGCTGGCAGCTCGAAACTTTGACCTGCAGAAATCCGAAGACATGCTCCGA AGGCACATGGAGTTCGGAAGCAACAAGACCTGGACAACATTGTACATGGCAGCCCCCTGAGGTGGTCATC CAGCTGTATGACTCGGGTGGTCTTTGTGGCTACGACTACGAAGGCTGCCCTGTGTACTTCAACATCATTGGG TCCTCGACCCCAAGGGTCTCCTGCTGTGAGCCTCCAAGCAGGATATGATCCGGAAGCGCATCAAAGTCTGT GAGCTGCTGTGTCATGAGTGTGAGCTGCAAACTCAGAAGCTGGGCAGGAAGATCGAGATGGCGCTGATGGTG TTTGACATGGAGGGGCTGAGCCTGAAACACCTGTGGAAGCCAGCTGTGGAGGTCTACCAGCAGTTTTTTAGC</p>

NOV29b

A disclosed NOV29b nucleic acid of 1305 nucleotides (also referred to as CG56187-03) encoding a CRAL-TRIO-like protein is shown in Table 29C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 14-16 and ending with a TGA codon at nucleotides 1262-1264. The start and stop codons are shown in bold in Table 29C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 29C. NOV29b nucleotide sequence (SEQ ID NO:123).

AGTTGACTGGTGGATGATGTGGGAAGGGTTAGGGGCGGGGTTGGTGGCCCCGAGGTCATGAGAGCTCCGCC
 GACCATCAGATCCTCCTCCGCTCAGTTCGGGAGAACCTCCAGGACCTGCTGCCATACCTGCCAATGCTGA
 TGACTACTTCCTCCTGCGCTGGCTGCGAGCTCGAAACTTTGACCTGCAGAAATCCGAAGACATGCTCCGAAG
 GCACATGGAGTTCGGGAAGCAACAAGACCTGGACAACATTGTACATGGCAGCCCCCTGAGGTCATCCAGCT
 GTATGACTCGGGTGGTCTTTGTGGCTACGACTACGAAGGCTGCCCTGTGTACTTCAACATCATTGGGTCCCT
 CGACCCCAAGGGTCTCCTGCTGTGAGCTCCAAGCAGGATATGATCCGGAAGCGCATCAAAGTCTGTGAGCT
 GCTGTTGCATGAGTGTGAGCTGCAAACTCAGAAGCTGGGCAGGAAGATCGAGATGGCGCTGATGGTGTGGA
 CATGGAGGGGCTGAGCCTGAAACACCTGTGGAAGCCAGCTGTGGAGGTCTACCAGCAGTTTCTTAGCATCCT
 GGAAGCAAATTATCCTGAGACCCTGAAGAATTAAATTGTTATTTCGAGCCCCAAACTGTTCCCGTGGCCTT
 CAACCTGGTCAAGTCGTTTCATGAGTGAGGAGACACGAGGAAGATTGTGATTCTGGGAGACAAGTGAAGCA
 GGAGCTGACAAAATTATCAGCCCCGACCAGCTGCCTGTGGAGTTTGGGGGACCATGACTGACCCCGATGG
 CCACCCCAAGTGCCTGACCAAGATCAACTATGGGGGTGAGGTGCCCAAGAGCTACTACCTGTGCGAGCAGGT
 GAGGCTGCAGTATGAGCACACGAGGTCCGTGGGCGCGGGCTCCTCCTGCAGGTGGAGAACGAGATCCTGTT
 CCCGGGCTGTGTGCTCAGGTGGCAGTTTGCTTCAGATGGTGGGGACATCGGCTTTGGGGTTTCTGTAAGAC
 CAAGATGGGGGAGCAGCAGAGTGCTAGGGAGATGACGGAGGTGCTGCCAGCCAGCGCTACAATGCCACAT
 GGTGCTGAGGATGGGAGCCTCACCTGCCTCCAGGCTGGCGTCTATGTCTGCGCTTCGACAACACCTACAG
 CCGGATGCATGCCAAGAAGCTCAGTACACTGTGGAGGTGTGCTTCCGACAAGGCTCTGAGGAGACGCT
 GCAGAGTCTCAAGGCGATGAGACCCTCCCCAACACAGTGAAGACCCAGCCACCTCCACCTGTGCACTCCAA
CCCCTTCAC

In a search of public sequence databases, the NOV29b nucleic acid sequence, located on chromosome 22, has 906 of 1212 bases (74%) identical to a gb:GENBANK-

10 ID:BC005759|acc:BC005759.1 mRNA from *Mus musculus* (*Mus musculus*, clone MGC:6302, mRNA, complete cds) ($E = 2.0e^{-137}$).

A disclosed NOV29b polypeptide (SEQ ID NO:124) encoded by SEQ ID NO:123 has 416 amino acid residues and is presented in Table 29D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV29b has no signal peptide and is
 15 likely to be localized extracellularly with a certainty of 0.4500. Alternatively, NOV29b may also localize to the mitochondrial membrane space with a certainty of 0.1000, to the lysosome (lumen) with a certainty of 0.1000, or to the microbody (peroxisome) with a certainty of 0.0779.

Table 29D. Encoded NOV29b protein sequence (SEQ ID NO:124).

MMWEGLGAGLVAPEVMRAPPTIRSSSAQFRENLDLPLILPNADDYFLLRWLRARNFDLQKSEDLRRHMEF
 RKQQLDLDNIWTPPEVIQLYDSGGLCGYDYEGCPVYFNIIIGSLDPKGLLSASKQDMIRKRIKVCELLLHE
 CELQTQKLGRKIEMALMVFDMEGLSLKHLWKPAVEVYQQFFSILEANYPETLKNLIVIRAPKLFPAFNLVK
 SFMSEETRRKIVILGDNWKQELTKFISPDQLPVEFGTMTDPDGHKCLTKINYGGEVPKSYLCEQVRLQY
 EHTRSVGRGSSLQVENEILFPGCVLRWQFASDGGDIFGFGVFLKTKMGEQQSAREMTEVLPSQRYNAHMPED

GSLTCLQAGVYVLRFDNTYSRMHAKKLSYTVLEVLLPDKASEETLQSLKAMRPSPTQ

A search of sequence databases reveals that the NOV29b amino acid sequence has 906 of 1212 amino acid residues (74%) identical to, and 906 of 1212 amino acid residues (74%) similar to, the 2529 amino acid residue gb:GENBANK-ID:BC005759|acc:BC005759.1 protein from *Mus musculus* (*Mus musculus*, clone MGC:6302, mRNA, complete cds) ($E = 2.0e^{-137}$).

NOV29b is predicted to be expressed in at least Bone, liver, brain, and prostate. . The sequence is predicted to be expressed in bone because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:BC005759|acc:BC005759.1) a closely related *Mus musculus*, clone MGC:6302, mRNA, complete cds homolog.

NOV29c

A disclosed NOV29c nucleic acid of 1218 nucleotides (also referred to as CG56189-01) encoding a CRAL-TRIO-like protein is shown in Table 29E. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 1-3 and ending with a TAG codon at nucleotides 1216-1218. The start and stop codons are shown in bold in Table 29E, and the 5' and 3' untranslated regions, if any, are underlined.

Table 29E. NOV29c nucleotide sequence (SEQ ID NO:125).

ATGTTCCGGGAGAACATCCAAGATGTGCTATCTGCGCTGCCAATCCTGATGACTACTTCTCTCGCGCTGGCTCCAAGCTCGGAGCTTTGACCTGCAGAAATCAGAGGACATGCTGAGGAAGCATATGGAGTTCCGGAAGCAA
CAAGACCTGGCCAACATCCTTGCCCTGGCAGCCCCAGAGGTGGTCAGGCTGTACAACGCTAACGGCATATGC
GGCCACGACGGTGAGGGCAGCCCTGTCTGGTACCACATTGTGGGAAGCCTGGACCCAAAGGCCTCTTGCTC
TCAGCCTCCAACAGGAGTTGCTCAGGGACAGCTTCCGGAGCTGCGAGCTGCTCCTCGGGAGTGTGAGCTG
CAGAGTCAGAAGCTGGGGAAGAGGGTGGAGAAAATCATAGCTATTTTGGTCTCGAAGGGCTGGGCTGAGG
GATCTGTGGAAGCCAGGAATAGAGCTTCTCCAGGAGTTTTCTCAGCACTTGAAGCAAATTACCTGAGATC
TTGAAGAGTTTAATTGTTGTGAGAGCCCCAAGCTATTCGCCGTAGCCTTCAACCTGGTCAAGTCTTACATG
AGTGAAGAGACACGCAGGAAGGTGGTGATTCTCGGAGATCTGATGGTTCTGCAATCCGAAGGTGTAGGGCAC
CCAACTGGTGTGAGGGCCCACTGCCTGGTGGGCTGCCAGACAAGTGAAGCAGGAGCTGACAAAATTCATC
AGCCCGACAGCTGCCCGTGGAGTTTGGGGGGACCATGACTGACCCCGATGGCAACCCCAAGTGCCTGACC
AAGATCAACTACGGGGGTGAGGTGCCAAGAGCTACTACCTGTGCAAGCAGGTGAGGCTGCAGTATGAGCAC
ACGAGTCCGTGGGCCGCGCTCCTCCTGCAAGTGGAGAACGAGATCCTGTTCCCGGCTGTGTGCTCAGG
TGGCAGTTTGTCTCAGATGGTGGGACATTGGCTTTGGGGTTTCTGAAGACCAAGATGGGGGAGCGGCAG
AGGGCTAGGGAGATGACAGAGGTGCTGCCAGCCAGCGCTACAATGCCACATGGTGCCTGAAGATGGGATT
CTCACCTGCCTCCAGGCCGGCAGCTATGTCTGAGGTTTTACAACACCTACAGCCTGGTTTCATTTCTAAACGC
ATCAGCTACACCGTGGAGGTACTGCTCCAGACCAAACTTCATGGAGAAGATGGAGAAATCTAG

In a search of public sequence databases, the NOV29c nucleic acid sequence, located on chromosome 22, has 418 of 532 bases (78%) identical to a gb:GENBANK-ID:HS130H16A|acc:AL096881.1 mRNA from *Homo sapiens* (Novel human mRNA similar to *Rattus norvegicus* 45 kDa secretory protein, AJ132352) ($E = 4.9e^{-129}$).

The disclosed NOV29c polypeptide (SEQ ID NO:126) encoded by SEQ ID NO:125 has 405 amino acid residues and is presented in Table 29F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV29c has no signal peptide and is likely to be localized extracellularly with a certainty of 0.4500. Alternatively, NOV29c

may also localize to the microbody (peroxisome) with a certainty of 0.2010, to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 29F. Encoded NOV29c protein sequence (SEQ ID NO:126).

MFRENIQDVLSALPNPDDYFLLRWLQARSFDLQKSEDMRLRKHMEFRKQQDLANILAWQPPEVVRLYNANGIC
GHDGEGSPVWYHIVGSLDPKGLLSASKQELLRDSFRSCCELLLRECELQSOKLGKRVEKIIAIFGLEGLGLR
DLWKPGIELLQEFFSALEANYPEILKSLIVVRAPKLFVAFNLVKSYSMEETRRKVVLGDLMPVASEGVGH
PTGVEGPLPGGLPDNWKQELTKFISPDQLPVEFGGTMTDPDGNPKCLTKINYGGEVPKSYLLCKQVRLQYEH
TRSVGRGSSLQVENEILFPGCVLRWQFASDGGDIGFGVFLKTKMGERQRAREMTEVLPSQRYNAHMPVEDGI
LTCLQAGSYVLRFYNTYSLVHSKRISYTVLEVLLPDQTFMEKMEKF

A search of sequence databases reveals that the NOV29c amino acid sequence has 157 of 176 amino acid residues (89%) identical to, and 166 of 176 amino acid residues (94%) similar to, the 406 amino acid residue ptrn:SPTREMBL-ACC:Q9UDX3 protein from *Homo sapiens* (Human) (WUGSC:H_DJ0539M06.4 PROTEIN) ($E = 2.6e^{-167}$).

NOV29c is predicted to be expressed in at least Bone, liver, brain, and prostate . This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in Bone because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:RNO132352|acc: AJ132352.1) a closely related *Rattus norvegicus* mRNA for 45 kDa secretory protein, partial homolog.

The disclosed NOV29a polypeptide has homology to the amino acid sequences shown in the BLASTP data listed in Table 29G.

Table 29G. BLAST results for NOV29a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 6624133 gb AAF19 259.1 AC004832_4 (AC004832)	similar to 45 kDa secretory protein [<i>Rattus norvegicus</i>]; similar to CAA10644.1 (PID:g4164418) [<i>Homo sapiens</i>]	406	387/398 (97%)	390/398 (97%)	0.0
gi 7110715 ref NP_0 36561.1 (NM_012429)	SEC14 (S. <i>cerevisiae</i>)-like 2; tocopherol- associated protein [<i>Homo sapiens</i>]	403	269/394 (68%)	331/394 (83%)	e-165

gi 16758646 ref NP_446253.1 (NM_053801)	SEC14 (S. cerevisiae)-like 2 [Rattus norvegicus]	403	271/394 (68%)	329/394 (82%)	e-164
gi 13543184 gb AAH05759.1 AAH05759 (BC005759)	Unknown (protein for MGC:6302) [Mus musculus]	403	273/394 (69%)	328/394 (82%)	e-164
gi 4164418 emb CAA10644.1 (AJ132352)	45 kDa secretory protein [Rattus norvegicus]	400	267/384 (69%)	326/384 (84%)	e-163

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 29H. In the ClustalW alignment of the NOV29 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 29H. ClustalW Analysis of NOV29

1) Novel NOV29a	(SEQ ID NO:122)
2) Novel NOV29b	(SEQ ID NO:124)
3) Novel NOV29c	(SEQ ID NO:126)
4) gi 6624133 gb AAF19259.1 AC004832_4 (AC004832) similar to 45 kDa secretory protein [Rattus norvegicus]; similar to CAA10644.1 (PID:gi4164418) [Homo sapiens] (SEQ ID NO:447)	
5) gi 7110715 ref NP_036561.1 (NM_012429) SEC14 (S. cerevisiae)-like 2; tocopherol-associated protein [Homo sapiens] (SEQ ID NO:448)	
6) gi 16758646 ref NP_446253.1 (NM_053801) SEC14 (S. cerevisiae)-like 2 [Rattus norvegicus] (SEQ ID NO:449)	
7) gi 13543184 gb AAH05759.1 AAH05759 (BC005759) Unknown (protein for MGC:6302) [Mus musculus] (SEQ ID NO:450)	
8) gi 4164418 emb CAA10644.1 (AJ132352) 45 kDa secretory protein [Rattus norvegicus] (SEQ ID NO:451)	

		10	20	30	40	50	60																									
NOV29a	1	MMW	EG	LG	AG	LV	AP	EV	MR	AP	PT	IR	SS	AA	CF	RE	NI	OD	LL	PI	LP	NA	DD	YF	LL	RW	LR	AR	NF	DL	Q	60
NOV29b	1	MMW	EG	LG	AG	LV	AP	EV	MR	AP	PT	IR	SS	AA	CF	RE	NI	OD	LL	PI	LP	NA	DD	YF	LL	RW	LR	AR	NF	DL	Q	60
NOV29c	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	33
gi 6624133	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50
gi 7110715	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50
gi 16758646	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50
gi 13543184	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50
gi 4164418	1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	50

		70	80	90	100	110	120																																																				
NOV29a	61	KSE	D	M	L	R	R	H	M	E	F	R	K	Q	D	I	D	N	I	V	I	W	Q	P	P	E	V	V	I	Q	L	Y	D	S	G	G	I	C	G	Y	D	Y	E	G	C	P	V	Y	F	N	I	I	G	S	L	D	P	K	120
NOV29b	61	KSE	D	M	L	R	R	H	M	E	F	R	K	Q	D	I	D	N	I	V	I	W	Q	P	P	E	V	V	I	Q	L	Y	D	S	G	G	I	C	G	Y	D	Y	E	G	C	P	V	Y	F	N	I	I	G	S	L	D	P	K	119
NOV29c	34	KSE	D	M	L	R	R	H	M	E	F	R	K	Q	D	I	D	N	I	V	I	W	Q	P	P	E	V	V	I	Q	L	Y	D	S	G	G	I	C	G	Y	D	Y	E	G	C	P	V	Y	F	N	I	I	G	S	L	D	P	K	92
gi 6624133	51	KSE	D	M	L	R	R	H	M	E	F	R	K	Q	D	I	D	N	I	V	I	W	Q	P	P	E	V	V	I	Q	L	Y	D	S	G	G	I	C	G	Y	D	Y	E	G	C	P	V	Y	F	N	I	I	G	S	L	D	P	K	109
gi 7110715	51	KSE	A	M	L	R	K	H	V	E	F	R	K	Q	D	I	D	K	I	I	S	W	Q	P	P	E	V	V	I	Q	Q	Y	L	S	G	G	M	C	G	Y	D	L	D	G	C	P	V	W	Y	D	I	I	G	P	L	D	A	K	109
gi 16758646	51	KSE	A	M	L	R	K	H	V	E	F	R	K	Q	D	I	D	K	I	I	S	W	Q	P	P	E	V	V	I	Q	Q	Y	L	S	G	G	R	C	G	Y	D	L	D	G	C	P	V	W	Y	D	I	I	G	P	L	D	A	K	109
gi 13543184	51	KSE	A	M	L	R	K	H	V	E	F	R	K	Q	D	I	D	K	I	I	S	W	Q	P	P	E	V	V	I	Q	Q	Y	L	S	G	G	R	C	G	Y	D	L	D	G	C	P	V	W	Y	D	I	I	G	P	L	D	A	K	109
gi 4164418	51	KSE	A	M	L	R	K	H	M	E	F	R	K	I	M	D	I	D	H	I	I	W	Q	P	P	E	V	V	I	Q	K	Y	M	P	S	G	I	C	G	Y	D	R	D	G	C	P	V	W	Y	D	I	I	G	P	L	D	P	K	109

			130	140	150	160	170	180			
	NOV29a	121	GLLLSASKQDMIRKRIKVC	ELLHECELO	TOKLGRKIE	MAIMVFD	MEGLSLKHLW	KPAVE	180		
5	NOV29b	120	GLLLSASKQDMIRKRIKVC	ELLHECELO	TOKLGRKIE	MAIMVFD	MEGLSLKHLW	KPAVE	179		
	NOV29c	93	GLLLSASKQELLRDSFR	SCELLRECELO	SOKLGKRV	EKIIAIFGL	EGLGLRDL	LWKPGE	152		
	gi 6624133	110	GLLLSASKQDMIRKRIKVC	ELLHECELO	TOKLGRKIE	MAIMVFD	MEGLSLKHLW	KPAVE	169		
	gi 7110715	110	GLLFSASKQDLLRTKMR	ECCELLLOE	CAHOTTKL	GRKVIET	ITITTYD	CEGLGLKHLW	KPAVE	169	
	gi 16758646	110	GLLFSASKQDLLRTKMR	DCCELLLOE	CTQQTAKL	GKKIET	ITITMIYD	CEGLGLKHLW	KPAVE	169	
10	gi 13543184	110	GLLFSASKQDLLRTKMR	DCCELLLOE	CIQQTAKL	GKKIET	ITITMIYD	CEGLGLKHLW	KPAVE	169	
	gi 4164418	110	GLLFSVTKQDLLRTKMR	DCERILHEC	DLQTERL	GRKIE	ITIVMIFD	CEGLGLKHFW	KPLVE	169	
			190	200	210	220	230	240			
15	NOV29a	181	VYQOFFSILEANYPETL	KNLIVIRAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	232		
	NOV29b	180	VYQOFFSILEANYPETL	KNLIVIRAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	231		
	NOV29c	153	LLOEFFSILEANYPETL	KSILIVIRAP	KLPFAV	NLVKSY	FMSEETR	RRKIVILG	212		
	gi 6624133	170	VYQOFFSILEANYPETL	KNLIVIRAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	221		
	gi 7110715	170	AYGEFLCMFEENYPETL	KRLFVVKAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	221		
20	gi 16758646	170	AYGEFLTMFEENYPETL	KRLFVVKAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	221		
	gi 13543184	170	AYGEFLTMFEENYPETL	KRLFVVKAP	KLPFAV	NLVKSE	FMSEETR	RRKIVILG	221		
	gi 4164418	170	VYQEFFGLLEENYPETL	KFLIVKAT	KLPFAV	GYNLMK	FMSEETR	RRKIVILG	221		
			250	260	270	280	290	300			
25	NOV29a	232	-----	DNWKQELTKF	ISPDL	PVEFG	TMTDP	DGNPKCLTK	INYGGEV	275	
	NOV29b	231	-----	DNWKQELTKF	ISPDL	PVEFG	TMTDP	DGNPKCLTK	INYGGEV	274	
	NOV29c	213	GVGHPTGVEGPLPGGL	PDNWKQELTKF	ISPDL	PVEFG	TMTDP	DGNPKCLTK	INYGGEV	272	
	gi 6624133	221	-----	DNWKQELTKF	ISPDL	PVEFG	TMTDP	DGNPKCLTK	INYGGEV	264	
30	gi 7110715	221	-----	ANWKEVLLKH	ISPDL	PVEYGG	TMTDP	DGNPKCKSK	INYGGEV	264	
	gi 16758646	221	-----	ANWKEVLLKH	ISPDL	PVEYGG	TMTDP	DGNPKCKSK	INYGGEV	264	
	gi 13543184	221	-----	ANWKEVLLKH	ISPDL	PVEYGG	TMTDP	DGNPKCKSK	INYGGEV	264	
	gi 4164418	221	-----	NSWKEGLLKL	ISPDL	PAHFG	GLTDP	DGNPKCLTK	INYGGEV	264	
			310	320	330	340	350	360			
35	NOV29a	276	PKSYLLCEQVRLQYEH	TRSVGRGSS	LQVENEIL	FPGCVLR	WQFASD	GGDIGFGV	FLKTKM	335	
	NOV29b	275	PKSYLLCEQVRLQYEH	TRSVGRGSS	LQVENEIL	FPGCVLR	WQFASD	GGDIGFGV	FLKTKM	334	
	NOV29c	273	PKSYLLCEQVRLQYEH	TRSVGRGSS	LQVENEIL	FPGCVLR	WQFASD	GGDIGFGV	FLKTKM	332	
40	gi 6624133	265	PKSYLLCEQVRLQYEH	TRSVGRGSS	LQVENEIL	FPGCVLR	WQFASD	GGDIGFGV	FLKTKM	324	
	gi 7110715	265	PRKYVVRDQVKQYEH	SVQISRGSS	HQVEY	EILFPG	CVLRWQ	FMSD	GADVGFCIF	FLKTKM	324
	gi 16758646	265	PKQYVVRDQVKQYEH	SVQISRGSS	HQVEY	EILFPG	CVLRWQ	FMSD	GADVGFCIF	FLKTKM	324
	gi 13543184	265	PKQYVVRDQVKQYEH	TVQISRGSS	HQVEY	EILFPG	CVLRWQ	FMSD	GADVGFCIF	FLKTKM	324
	gi 4164418	265	PKSMVVRDQVKQYEH	SVQISRGSS	HQVEY	EILFPG	CVLRWQ	FMSD	GADVGFGV	FLKTKM	324
45			370	380	390	400	410	420			
50	NOV29a	336	GEQOSAREMTEVLPS	QRYNAH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SRMHAKKLSY	TVEVL	393
	NOV29b	335	GEQOSAREMTEVLPS	QRYNAH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SRMHAKKLSY	TVEVL	394
	NOV29c	333	GERORAREMTEVLPS	QRYNAH	MVPEDG	SLTCLQ	AGSYVLR	FDNTYS	SLVHAKKLSY	TVEVL	392
	gi 6624133	325	GEQOSAREMTEVLPS	QRYNAH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SRMHAKKLSY	TVEVL	384
	gi 7110715	325	GERORAREMTEVLPS	QRYNSH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SFIHAKKVNFT	TVEVL	384
	gi 16758646	325	GERORAREMTEVLPS	QRYNSH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SFIHAKKVNFT	TVEVL	384
	gi 13543184	325	GERORAREMTEVLPS	QRYNSH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SFIHAKKVNFT	TVEVL	384
	gi 4164418	325	GERORAREMTEVLPS	QRYNAH	MVPEDG	SLTCLQ	AGVYVLR	FDNTYS	SFIHAKKVNFT	TVEVL	384
55			430	440							
60	NOV29a	394	LPDKASEETLQSLK	AMRPSPTQ						415	
	NOV29b	395	LPDKASEETLQSLK	AMRPSPTQ						416	
	NOV29c	393	LPDQTFMEKMEK	-----						405	
	gi 6624133	385	LPDKASEETLQSLK	AMRPSPTQ						406	
	gi 7110715	385	LPDKASEEKKQCL	CAGTEPK	---					403	
	gi 16758646	385	LPDKAAEEKLNQ	QCAVTPK	---					403	
	gi 13543184	385	LPDKAAEEKMNQ	QCAVTPK	---					403	
65	gi 4164418	385	LPDEGMOKYDEEL	---TPI---						400	

Tables 29I-J list the domain descriptions from DOMAIN analysis results against NOV29. This indicates that the NOV29 sequence has properties similar to those of other proteins known to contain this domain.

Table 29I Domain Analysis of NOV29

gnl|Smart|smart00516, SEC14, Domain in homologues of a *S. cerevisiae* phosphatidylinositol transfer protein (Sec14p); Domain in homologues of a *S. cerevisiae* phosphatidylinositol transfer protein (Sec14p) and in RhoGAPs, RhoGEFs and the RasGEF, neurofibromin (NF1). Lipid-binding domain. The SEC14 domain of Dbp is known to associate with G protein beta/gamma subunits. (SEQ ID NO:822)
CD-Length = 157 residues, 96.8% aligned
Score = 131 bits (329), Expect = 9e-32

5	NOV29:	90	VIQLYDSGGLCGYDYEGCPVYFNIIGSLDPKGLLSASKQDMIRKRIKVCCELLLHECELQ	149
			+ + + + + + +	
	Sbjct:	4	VGKAYIPGGR--YDKDGRPVLVFRAGRFDLK---SVTLEELLRYLVVYVLEKALQE----	53
10	NOV29:	150	TQKLGRKIEALMVFDMEGLSLKHLWKPAVEVYQQFFSILEANYPETLKNLIVIRAPKLF	209
			+ + + + + + + + + + +	
	Sbjct:	54	-EKKTGGIEGFTTIFDLKGLSMSN--PDLGVLRKILKILQDHYPERLGKVYIINPPWFF	109
15	NOV29:	210	PVAFNLVKSFMSEETRRKIVILGDNWKQELTKFISPDQLPVEFGGT	255
			+ + + + + + + + + + +	
	Sbjct:	110	RVLWKIIPFLSEKTRKIRFVGPDSEELLEYIDPEQLPEELGGT	155

Table 29J Domain Analysis of NOV29

gnl|Pfam|pfam00650, CRAL TRIO, CRAL/TRIO domain.. The original profile has been extended to include the carboxyl domain from the known structure of Sec14. (SEQ ID NO:823)
CD-Length = 185 residues, 98.9% aligned
Score = 120 bits (300), Expect = 2e-28

20	NOV29:	73	RKQDLNIV-TWQPPEVVIQLYDSGGLCGYDYEGCPVYFNIIGSLDPKGLLSASKQDM	131
			+ + + + + + + + + +	
	Sbjct:	3	RREFGVDTILEEATYPKEVIKALYPQFIHGSDKDGPRVYLERRGQLNLKMLFITTVERM	62
25	NOV29:	132	IRKRIKVCE-LLLHECELQTQKLGRKIEALMVFDMEGLSL-KHLWKPAVEVYQQFFSIL	189
			+ + + + + + + + + + + + +	
	Sbjct:	63	VRNLVYEMEQALLYLLPACSRKVGTLINGSCTVFDLKGVSVSANWVPGVL--KKVLNIL	120
30	NOV29:	190	EANYPETLKNLIVIRAPKLFPAFNLVKSFMSEETRRKIVILGDNWKQELTKFISPDQLP	249
			+ + + + + + + + + + +	
	Sbjct:	121	QDYYPRLGKFYILINAPWLFSTVYKLIKPLDPKTRKIFVLGNY-KSELLQYIPADNLP	179
	NOV29:	250	VEFGGT	255
			+	
	Sbjct:	180	AKLGGT	185

35 Vitamin E (alpha-tocopherol) is an essential dietary nutrient for humans and animals. The mechanisms involved in cellular regulation as well as in the preferential cellular and tissue accumulation of alpha-tocopherol are not yet well established. We previously reported (Stocker, A., Zimmer, S., Spycher, S. E., and Azzi, A. (1999) IUBMB Life 48, 49-55) the

identification of a novel 46-kDa tocopherol-associated protein (TAP) in the cytosol of bovine liver. Here, we describe the identification, the molecular cloning into *Escherichia coli*, and the in vitro expression of the human homologue of bovine TAP, hTAP. This protein appears to belong to a family of hydrophobic ligand binding proteins, which have the CRAL (cis-retinal binding motif) sequence in common. By using a biotinylated alpha-tocopherol derivative and the IASys resonant mirror biosensor, the purified recombinant protein was shown to bind tocopherol at a specific binding site with $K(d) 4.6 \times 10(-7) m$. Northern analyses showed that hTAP mRNA has a size of approximately 2800 base pairs and is ubiquitously expressed. The highest amounts of hTAP message are found in liver, brain, and prostate. In conclusion, hTAP has sequence homology to proteins containing the CRAL_TRIO structural motif. TAP binds to alpha-tocopherol and biotinylated tocopherol, suggesting the existence of a hydrophobic pocket, possibly analogous to that of SEC14. Zimmer S. et al. (J Biol Chem. 2000 Aug 18;275(33):25672-80.)

The disclosed NOV29 nucleic acid of the invention encoding a CRAL-TRIO -like protein includes the nucleic acid whose sequence is provided in Table 29A, 29C, 29E or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 29A, 29C, or 29E while still encoding a protein that maintains its CRAL-TRIO -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 26 percent of the bases may be so changed.

The disclosed NOV29 protein of the invention includes the CRAL-TRIO -like protein whose sequence is provided in Table 29B, 29D, or 29F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 29B, 29D, or 29F while still encoding a protein that maintains its CRAL-TRIO -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this CRAL-TRIO -like protein (NOV29) is a member of a "CRAL-TRIO family". Therefore, the NOV29 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV29 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis colon cancer, leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency and cancer; and prostate disorders including prostate cancer, and/or other diseases and pathologies.

NOV29 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV29 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV29 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV30

A disclosed NOV30 nucleic acid of 717 nucleotides (also referred to as CG56191-01) encoding a novel Ryudocan-like protein is shown in Table 30A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 22-24 and ending with a TAG codon at nucleotides 658-660. Putative untranslated regions, if any, found upstream from

the initiation codon and downstream from the termination codon are underlined in Table 30A, and the start and stop codons are in bold letters.

Table 30A. NOV30 Nucleotide Sequence (SEQ ID NO:127)	
CAGGCTGTT	<u>CACCTCTCTGGATGGCGGTACCCACTGCCCCGCCCTCCTGCTCCTGCTGCTGCTGCTGCT</u>
TTTTCAGGCACCC	<u>CACCCCTGAGTCAATCCAAGAACTGAGGTCATCAACCCAGGACCGCCTAGGG</u>
GCCCAAAC	<u>TCTCCAGATCCCTACTGGAAGACTCTGGATGTGGGTGTGGGGG</u>
GAGCTCTCTGGCTCTAGAGATATTGATGAGTCAAGGGACCCCAAGATCATCCCTGAAGTGATCCAACCCCTT	
GGTGCTTCTAGATAACCA	<u>CATCCCTGAGAGGGCAGGGCCTGGGAACCTGGTCCCCACTGAAACCAAGGAAC</u>
TGGAGGACAACGAGGTCATCCCAGGAGGATCTCACTCTCTGCGGGGGACAGGATGTGTCCAATAAGGCA	
CCCATGTCCAACACTGCC	<u>CAGGGCAGCAACATCTTTGAGAGAATGGAGGTCGTGGCAGTCTGATTGTGGA</u>
CAGCATCGCGGGCATCCTCTCTGCTGTTTTCTGATCCTGCTTCTGGTGAACCATATGAAGAAGGATGAAG	
GCAGAAACGACCTGAGCAGGAAGCCCATCTACAAAAAAGCCCCTAGCAAGGAGTTATTACGCTTCTTCTAT	
GAGCACTGGTTTGACTT	<u>TAGGGGATAGGGAAGTCGGAGGATTTGCAGAGTGGCCATTAGGATGCGGGAG</u>
GACAACC	

The NOV30 nucleic acid was identified on chromosome 22 and has 553 of 708 bases (78%) identical to a gb:GENBANK-ID:HUMRYUDO|acc:D13292.1 mRNA from *Homo sapiens* (mRNA for ryudocan core protein) ($E = 2.2e^{-82}$).

A disclosed NOV30 polypeptide (SEQ ID NO:128) encoded by SEQ ID NO:127 is 212 amino acid residues and is presented using the one-letter code in Table 30B. Signal P, Psort and/or Hydropathy results predict that NOV30 contains a signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.4600. The most likely cleavage site for NOV30 is between positions 23 and 24: TPT-TP.

Table 30B. Encoded NOV30 protein sequence (SEQ ID NO:128)	
MAVPTAPALLLLLLLLFAGTPTTPESIQETEVINPGPPRGNFNSRSLLEDSCGCGWQEPDDSELSGSRDI	
DES RDPKIIPEVIQPLVLLDNHIIERAGPGNLVPTETKELEDNEVI PRRI SLSAGDQDVSNKAPMSNTAQGS	
NIFERMEVVAVLIVDSIAGILSAVFLILLLVNHHMKKDEGRNDLSRKPIYKKAPSKELLRFFFEYHWFGL	

The disclosed NOV30 amino acid sequence has 121 of 198 amino acid residues (61%) identical to, and 140 of 198 amino acid residues (70%) similar to, the 202 amino acid residue ptnr:SWISSPROT-ACC: P34901 protein from *Rattus norvegicus* (Rat) (Syndecan-4 Precursor (Ryudocan Core Protein)) ($E = 1.9e^{-51}$).

NOV30 is predicted to be expressed in at least myeloid tissue, B-cell lymphoma, including B-cell precursor lymphoblastic leukemia, lymphoplasmacytoid, immunoblastic, lymphocytic/CLL, hairy cell leukemia, large B-cell, mantle-cell, marginal zone and follicular, lymphomas, endothelia, Lymphopoietic and bone marrow (BM) plasma cells (PCs). This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV30 is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HUMRYUDO|acc: D13292.1) a

closely related Human mRNA for ryudocan core protein homolog in species *Homo sapiens*: myeloid tissue.

NOV30 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 30C.

5

Table 30C. BLAST results for NOV30					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14771140 ref XP_009530.3 (XM_009530)	syndecan 4 (amphiglycan, ryudocan) [<i>Homo sapiens</i>]	198	119/197 (60%)	136/197 (68%)	9e-49
gi 4506861 ref NP_02990.1 (NM_002999)	syndecan 4 (amphiglycan, ryudocan) [<i>Homo sapiens</i>]	198	120/197 (60%)	137/197 (68%)	2e-45
gi 6981522 ref NP_036781.1 (NM_012649)	ryudocan/syndecan 4 [<i>Rattus norvegicus</i>]	202	119/199 (59%)	139/199 (69%)	3e-45
gi 6755442 ref NP_035651.1 (NM_011521)	syndecan 4 [<i>Mus musculus</i>]	198	117/199 (58%)	136/199 (67%)	6e-41
gi 1351051 sp P49416 SDC4_CHICK	SYNDECAN-4 PRECURSOR	197	80/216 (37%)	105/216 (48%)	1e-14

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 30D.

Table 30D Clustal W Sequence Alignment

10

1) NOV30 (SEQ ID NO:128)
2) gi|14771140|ref|XP_009530.3| (XM_009530) syndecan 4 (amphiglycan, ryudocan) [*Homo sapiens*] (SEQ ID NO:452)
3) gi|4506861|ref|NP_002990.1| (NM_002999) syndecan 4 (amphiglycan, ryudocan) [*Homo sapiens*] (SEQ ID NO:453)
4) gi|6981522|ref|NP_036781.1| (NM_012649) ryudocan/syndecan 4 [*Rattus norvegicus*] (SEQ ID NO:454)
5) gi|6755442|ref|NP_035651.1| (NM_011521) syndecan 4 [*Mus musculus*] (SEQ ID NO:455)
6) gi|1351051|sp|P49416|SDC4_CHICK SYNDECAN-4 PRECURSOR (SEQ ID NO:456)

15

20

			10	20	30	40	50	60											
NOV30	1	MAVPTAPAL	LLLLLLLL	FAGTP	--TTPESIQ	ETEVINP	-GPPRC	PNFSS	RLLED	SGCCW	57								
gi 14771140	1	----	MAPARLFA	-----	VAESIRE	TEVIDP	QDLLEG	RYFSGAL	PDD	EDVVGP	51								
gi 4506861	1	----	MAPARLFA	-----	VAESIRE	TEVIDP	QDLLEG	RYFSGAL	PDD	EDVVGP	51								
gi 6981522	1	----	MAPVCLFA	LLLLLL	GGFPVAP	GESIRE	TEVIDP	QDLLEG	RYFSGAL	PDD	EDAGGL	56							
gi 6755442	1	----	MAPACLLA	LLLLLL	GGFPVAP	GESIRE	TEVIDP	QDLLEG	RYFSGAL	PDD	EDAGGS	56							
gi 1351051	1	--MPT	PRAAFL	LGLLL	AAAAA	-----	ESVRE	TETMD	ARWLD	NVNG	--SGDL	PDD	EDIGEF	50					
			70	80	90	100	110	120											
NOV30	58	GQ	--EP	DDSD	-----	ELSGS	RDID	ESRDP	KIIP	EVTC	PLVLL	DNH	IPER	AGSG	NLVPT	ETIKEL	112		
gi 14771140	52	GQ	--ES	DDDF	-----	ELSGS	GDLD	DDLED	SMIG	PEV	VHPL	VPLD	NH	IPER	AGSG	SQVPT	EPKKL	106	
gi 4506861	52	GQ	--ES	DDDF	-----	ELSGS	GDLD	DDLED	SMIG	PEV	VHPL	VPLD	NH	IPER	AGSG	SQVPT	EPKKL	106	
gi 6981522	57	EQ	--DS	DF	-----	ELSGS	GDLD	DDTE	EPRT	FF	EVIS	PLVPL	DNH	IPEN	AQPG	IRVPS	EPKEL	110	
gi 6755442	57	D	-----	DF	-----	ELSGS	GDLD	DDTE	EPRT	FF	EVIS	PLVPL	DNH	IPEN	AQPG	IRVPS	EPKEL	107	
gi 1351051	51	TPHLT	SDEF	DIDDT	SGSG	GYSD	YD	DAI	YLT	TVD	TPAT	S	DNM	IP	GD	TER	KMEGE	KKNTM	108

35

15

20

gnl|Pfam|pfam01034, Syndecan, Syndecan domain. Syndecans are transmembrane heparin sulfate proteoglycans which are implicated in the binding of extracellular matrix components and growth factors (SEQ ID NO:824)
CD-Length = 359 residues, 21.7% aligned
Score = 41.6 bits (96), Expect = 5e-05

30

40

suggesting that proteoglycans of the heparan sulfate type are intimately involved. Consistently, inhibitors such as p-nitrophenyl-beta-d-xylopyranoside and chlorate interfering with heparan sulfate proteoglycan biosynthesis reduced the total number of kininogen binding sites in a time- and concentration-dependent manner (up to 67%). *In vitro* binding studies demonstrated that biotinylated H-kininogen binds to heparan sulfate glycosaminoglycans via domains D3 and D5(H) and that the presence of Zn(2+) promotes this association. Cloning and over-expression of the major endothelial heparan sulfate-type proteoglycans syndecan-1, syndecan-2, syndecan-4, and glypican in HEK293t cells significantly increased total heparan sulfate at the cell surface and thus the number of kininogen binding sites (up to 3.3-fold). This gain in kininogen binding capacity was completely abolished by treating transfected cells with heparinases. It was concluded that heparan sulfate proteoglycans on the surface of endothelial cells provide a platform for the local accumulation of kininogens on the vascular lining. This accumulation may allow the circumscribed release of short-lived kinins from their precursor molecules in close proximity to their sites of action (Renne et al., J Biol Chem 2000, 275(43):33688-96).

Lymphopoietic cells require interactions with bone marrow stroma for normal maturation and show changes in adhesion to matrix during their differentiation. Syndecan, a heparan sulfate-rich integral membrane proteoglycan, functions as a matrix receptor by binding cells to interstitial collagens, fibronectin, and thrombospondin. Therefore, it was asked whether syndecan was present on the surface of lymphopoietic cells. In bone marrow, syndecan was only found on precursor B cells. Expression changes with pre-B cell maturation in the marrow and with B-lymphocyte differentiation to plasma cells in interstitial matrices. Syndecan on B cell precursors is more heterogeneous and slightly larger than on plasma cells. Syndecan 1) is lost immediately before maturation and release of B lymphocytes into the circulation, 2) is absent on circulating and peripheral B lymphocytes, and 3) is reexpressed upon their differentiation into immobilized plasma cells. Thus, syndecan is expressed only when and where B lymphocytes associate with extracellular matrix. These results indicate that B cells differentiating *in vivo* alter their matrix receptor expression and suggest a role for syndecan in B cell stage-specific adhesion (Sanderson et al., Cell Regul 1989,1(1):27-35).

Detection of abnormal numbers and/or distribution of bone marrow (BM) plasma cells (PCs) on trephine biopsies can be important in the differential diagnosis of multiple myeloma (MM) and other PC disorders. A variety of immunohistochemical markers can potentially improve the specificity and sensitivity of PC detection on routine histological sections obtained from trephine BM biopsies, but most of them are not completely satisfactory. In one

study, the antibody CD138/B-B4, which is an optimal marker for PC detection on BM aspirates by flow cytometry, was investigated to determine whether it can be used successfully for the identification of PCs also on formalin-fixed, decalcified biopsies. A series of samples including normal BM, MM, monoclonal gammopathies of undetermined significance, and B-cell lymphoma of various types, including B-cell precursor lymphoblastic leukemia, lymphoplasmacytoid, immunoblastic, lymphocytic/CLL, hairy cell leukemia, large B-cell, mantle-cell, marginal zone and follicular lymphomas, have been investigated for CD138 expression using a sensitive immunohistochemical technique. Within the BM microenvironment, CD138 was characterized by excellent sensitivity and specificity. Virtually all normal and neoplastic PCs expressed clear-cut membrane CD138 immunostaining, whereas all other cell types did not. All cases of MM, including plasmablastic and leukemic cases, showed strong immunoreactivity. Conversely, all B-cell lymphomas, including all cases characterized by secretive features, lymphoplasmacytoid, and immunoblastic lymphomas, were completely negative. These results demonstrate that CD138 is a highly sensitive and specific marker that is useful for the rapid and precise localization of normal and neoplastic PCs on routine BM sections. In addition, because of its clear-cut cell membrane localization, CD138 can be used successfully in double-marker immunostaining reactions to evaluate precisely nuclear prognostic markers such as Ki67 and p53 in MMs (Chilosi et al., *Mod Pathol* 1999, 12(12):1101-6).

Monoclonal antibody therapy has emerged as a viable treatment option for patients with lymphoma and some leukemias. It is now beginning to be investigated for treatment of multiple myeloma. There are relatively few surface antigens on the plasma cells that are suitable for antibody-directed treatment. Possible molecules include HM1.24, CD38, ICAM-1 (CD54), CD40, CD45, CD20, and syndecan 1. There is now some clinical experience with anti-CD38 antibody in lymphoma and myeloma. However, to date, there has been minimal clinical activity observed. Additional antibodies are entering clinical trials. A new approach involves the generation of an anti-CD38 single-chain variable fragment (scFv) construct that acts as the carrier of a toxin gene instead of being conjugated directly to the toxin itself. It is hoped that expression of the toxin by CD38+ plasma cells will promote suicide of the malignant cells without affecting normal cells or generating an immunologic response to the toxin. Ongoing clinical trials are also attempting to target B-cell antigens such as CD20. Although CD20 is present only on 20% of myeloma cells, it may be present on myeloma precursor cells. This treatment has met with success in follicular lymphoma and is now being evaluated in clinical trials in both Europe and the United States for myeloma. Although these

clinical trials are in very early stages, researchers are beginning to understand that antibody therapy can be used not only as a carrier molecule of radioisotopes and toxins, but also as molecules that can trigger tumor cells and promote growth arrest or apoptosis (Maloney et al., Semin Hematol 1999, 36(1 Suppl 3):30-3).

5 The NOV30 nucleic acid of the invention encoding a Ryudocan-like protein includes the nucleic acid whose sequence is provided in Table 30A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 30A while still encoding a protein that maintains its Ryudocan-like activities and physiological functions, or a fragment of such a nucleic acid.

10 The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar

15 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 22% of the residues may be so changed.

20 The NOV30 protein of the invention includes the Ryudocan-like protein whose sequence is provided in Table 30B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 30B while still encoding a protein that maintains its Ryudocan-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 63 of the bases

25 may be so changed.

 The NOV30 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: brain disorders including epilepsy,

30 eating disorders, schizophrenia, ADD, cancer, heart disease, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders, psoriasis, colon cancer, leukemia, AIDS, thalamus disorders, metabolic disorders including diabetes and obesity, lung diseases such as asthma, myelomas, emphysema, cystic fibrosis, and cancer, pancreatic disorders

including pancreatic insufficiency and cancer, and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like.

NOV30 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. For example the disclosed NOV30 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV31

A disclosed NOV31 nucleic acid of 683 nucleotides (also referred to as CG56392-01) encoding a novel Sulfur-rich Keratin-like protein is shown in Table 31A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 46-48 and ending with a TGA codon at nucleotides 652-654. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 31A, and the start and stop codons are in bold letters.

Table 31A. NOV31 Nucleotide Sequence (SEQ ID NO:129)

GAGCTGTGTAAACAGCAACCGGAAAGAGAAACAATGGTGTGTTTCCT**ATG**TGGGATATAAAGAGCCGGGGCTC
AGGGGGCTCCACACCTGCACCTCCTTCTCACCTGCTCCTCTACCTGCTCCACCCTCAATCCACCAGAACCA
TGGGCTGCTGTGGCTGCTCCGGAGGCTGTGGCTCCAGCTGTGGAGGCTGTGACTCCAGCTGTGGGAGCTGT
GGCTCTGGCTGCAGGGGCTGTGGCCCCAGCTGCTGTGCACCCGCTCTACTGCTGCAAGCCCGTGTGCTGCTG
TGTTCAGCCTGTTTCCTGCTCTAGCTGTGGCAAGCGGGCTGTGGCTCCTGTGGGGCTCCAAGGGAGGCT
GTGGTTCTTGTGGCTGCTCCAGTGCAAGTGTGCAAGCCCTGCTGTTGCTCTTCAGGCTGTGGGTATCC
TGCTGCCAGTGCAAGTGTGCAAGCCCTACTGCTCCAGTCCAGCTGTTGTAAGCCCTGTTGCTGCTCCTC
AGGCTGTGGATCATCCTGCTGCCAGTCCAGCTGCTGCAAGCCCTGCTGCTGCCAGTCCAGCTGCTGTGTCC
CCGTGTGCTGCCAGTCCAGCTGCTGCAAGCCCTGTTGCTGCCAGTCCAACTGTTGTGTCCCTGTGTGCTGC
CAGTGTAAGATCT**GAGGCTCTAGTGGGAAACCTCAGGTAGCTCC**

The NOV31 nucleic acid was identified on chromosome 11 and has 654 of 683 bases (95%) identical to a gb:GENBANK-ID:HSA6693|acc:AJ006693.1 mRNA from *Homo sapiens* (UHS KerA gene) ($E = 3.3e^{-136}$).

A disclosed NOV31 polypeptide (SEQ ID NO:130) encoded by SEQ ID NO:129 is 202 amino acid residues and is presented using the one-letter code in Table 31B. Signal P, Psort and/or Hydropathy results predict that NOV31 contains a signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. The most likely cleavage site for a NOV31 peptide is between amino acids 32 and 33: TRT-MG.

Table 31B. Encoded NOV31 protein sequence (SEQ ID NO:130)

MWDIKSRGSGGSTPAPPShLLLYLLHPQSTRTMGCCGCGSGGCGSSCGGCDSSCGSCGSGCRGCGPSCCAPVY
CCKPVPCCCVACSCSSCGKRGCGSGGSKGGCGSCGCSQCSCCKPCCSSGCGSSCCQCSCKPYCSQSSCC
KPCSSSGCGSSCCQSSCKPCCQSSCCVPVCCQSSCKPCCQSNCCVPVCCQCKI

The disclosed NOV31 amino acid sequence has 158 of 170 amino acid residues (92%) identical to, and 158 of 170 amino acid residues (92%) similar to, the 169 amino acid residue ptnr:SWISSNEW-ACC:P26371 protein from *Homo sapiens* (Human) (Keratin, Ultra High-Sulfur Matrix Protein A (Uhs Keratin A) (Uhs Kera)) ($E = 1.8e^{-101}$).

NOV31 is predicted to be expressed in at least Kidney, Pancreas, Testis and Whole Organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV31 is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID:gb:GENBANK-ID:HSA6693|acc:AJ006693.1) a closely related *Homo sapiens* UHS Kera gene homolog in species *Homo sapiens*: Kidney, Pancreas and Testis.

NOV31 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 31C.

Table 31C. BLAST results for NOV31

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12835376 dbj BAB23238.1 (AK004258)	data source:SPTR, source key:Q64526, evidence:ISS-puta tive-similar to ULTRA-HIGH SULPHUR KERATIN [<i>Mus musculus</i>]	195	76/157 (48%)	91/157 (57%)	5e-13
gi 2136964 pir I46489	cysteine-rich hair keratin associated protein - rabbit	126	53/120 (44%)	72/120 (59%)	3e-11
gi 12844600 dbj BAB26426.1 (AK009665)	data source:SPTR, source key:Q28707, evidence:ISS-homo log to CYSTEINE RICH HAIR KERATIN ASSOCIATED PROTEIN-putative [<i>Mus musculus</i>]	168	59/116 (50%)	70/116 (59%)	1e-10
gi 15082220 ref NP_149048.1 (NM_033059)	keratin associated protein 4.14 [<i>Homo sapiens</i>]	195	56/122 (45%)	65/122 (52%)	2e-10

The major histological components of the hair follicle are the hair cortex and cuticle. The hair cuticle cells encase and protect the cortex and undergo a different developmental program to that of the cortex. In one study, the molecular characterization of a set of evolutionarily conserved hair genes which are transcribed in the hair cuticle late in follicle development was reported. Two genes were isolated and characterized, one expressed in the human follicle and one in the sheep follicle. Each gene encodes a small protein of 16 kD, containing greater than 50 cysteine residues, ranging from 31 to 36 mol% cysteine. Their high cysteine content and in vitro expression data identify them as ultra-high-sulfur (UHS) keratin proteins. The predicted proteins are composed almost entirely of cysteine-rich and glycine-rich repeats. Genomic blots reveal that the UHS keratin proteins are encoded by related multigene families in both the human and sheep genomes. Tissue in situ hybridization demonstrates that the expression of both genes is localized to the hair fiber cuticle and occurs at a late stage in fiber morphogenesis (MacKinnon et al., J Cell Biol 1990, 111(6 Pt 1):2587-600).

273

invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 31A while still encoding a protein that maintains its Sulfur-rich Keratin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 5% of the residues may be so changed.

The NOV31 protein of the invention includes the Sulfur-rich Keratin-like protein whose sequence is provided in Table 31B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 31B while still encoding a protein that maintains its Sulfur-rich Keratin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 56% of the bases may be so changed.

The NOV31 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: brain disorders including epilepsy, eating disorders, schizophrenia, ADD, cancer, heart disease, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders, psoriasis, colon cancer, leukemia, AIDS, thalamus disorders, metabolic disorders including diabetes and obesity, lung diseases such as asthma, emphysema, cystic fibrosis, and cancer, pancreatic disorders including pancreatic insufficiency and cancer, and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like.

NOV31 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-

NOVX Antibodies” section below. For example the disclosed NOV31 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV32

A disclosed NOV32 nucleic acid of 1575 nucleotides (also referred to as CG56686-01) encoding a novel DNMT1 associated protein-1 (DMAP)-like protein is shown in Table 32A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 94-96 and ending with a TGA codon at nucleotides 1573-1575. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 32A, and the start and stop codons are in bold letters.

Table 32A. NOV32 Nucleotide Sequence (SEQ ID NO:131)

CTTGGAGGCTGCAGGTCCGGACCCAGGTGCGGAAGTGCAGAGGGCCAGGCACTGACCCCTTGACCTCCGGTG
GCTCCCCCATCTCTCAGGCGCGATGGCTACGGGCGCGGATGTACGGGACATTCTAGAACTCGGGGGTCCAG
AAGGGGATGCAGCCTCTGGGACCATCAGCAAGAAGGACATTATCAACCCGGACAAGAAAAATCCAAGAAG
TCCTCTGAGACACTGACTTTCAAGAGGCCCGAGGGCATGCACCGGAAGTCTATGCCTTGCTCTACTCTGA
CAAGAACAAGGGCTCCTGCTTGCTTAGCAGGATGCAGGAGGACCTGAAGTCTTTTGCTCCAGGACATGACT
TTCTTGCTATAGGGGATGCACCCCCACTGTACCCAGTGACACTGGCCAGGGATACCGTACAGTGAAGGCC
AAGTTGGGCTCCAAGAAGGTGCGGCTTGGAAGTGGATGCCATTACCAACCCGGCCGCAAGGACGGAGC
AATGTTCTTCCACTGGCGACGTGCAGCGGAGGAGGGCAAGGACTACCCCTTTGCCAGGTTCAATAAGACTG
TGCAGGTGCTGTGTACTCGGAGCAGGAGTACCAGCTTATCTCCACGATGATGCTTGGACTAAGGCAGAA
ACTGACCACCTCTTTGACCTCAGCCGCGCTTTGACCTGCGTTTTGTTGTTATCCATGACCGGTATGACCA
CCAGCAGTTCAAGAAGCGTTCTGTGGAAGACCTGAAGGAGCGGTACTACCACATCTGTGCTAAGCTTGCCA
ACGTGCGGGCTGTGCCAGGCACAGACCTTAAGATACCAGTATTTGATGCTGGGCACGAACGACGGCGGAAG
GAACAGCTTGAGCGTCTCTACAACCGGACCCAGAGCAGGTGGCAGAGGAGGAGTACCTGCTACAGGAGCT
GCGCAAGATTGAGGCCCGGAAGAAGGAGCGGGAGAAACGCAGCCAGGACCTGCAGAAGCTGATCAGAGCGG
CAGACACCACTGCAGAGCAGCGCGCACGGAACGCAAGGCCCCCAAAAAGAAGTACCCAGAAAAAGGAG
GCTGAGAAGCCGGCTTCTCTGAGACTGCAGGCATCAAGTTTCCAGACTTCAAGTCTGCAGGTGTCACGCT
GCGGAGCCAACGGATGAAGCTGCCAAGCTCTGTGGGACAGAAGAAGATCAAGGCCCTGGAACAGATGCTGC
TGGAGCTTGGTGTGGAGCTGAGCCCGACACCTACGGAGGAGCTGGTGCACATGTTCAATGAGCTGCCAAGC
GACCTGGTGCTGCTCTACGAGCTCAAGCAGGCTGTGCCAAGTGCAGATATGAGCTGCAGATGCTGCGGCA
CCGTGATGAGGCACTGGCCCGGGCTGGTGTGCTAGGGGGCCCTGCCACACCAGCATCAGGCCAGGCCCGG
CCTCTGCTGAGCCGGCAGTGACTGAACCCGCACTTGGTCTGACCCCAAGGACACCATCATTTGATGTGGTG
GGCGCACCCCTCACGCCCAATTGAGAAAGCGACGGAGTGGCCTCCAGTCTATCTCCGTGAAGAAAGC
CAAGAAGCCGTGA

The NOV32 nucleic acid was identified on chromosome 1p34 and has 1244 of 1273 bases (97%) identical to a gb:GENBANK-ID:AF265228|acc:AF265228.1 mRNA from *Homo sapiens* (DNMT1 associated protein-1 (DMAP1) mRNA, complete cds) ($E = 1.0e^{-309}$).

A disclosed NOV32 polypeptide (SEQ ID NO:132) encoded by SEQ ID NO:131 is 493 amino acid residues and is presented using the one-letter code in Table 32B. Signal P, Psort and/or Hydropathy results predict that NOV32 does not contain a signal peptide and is likely to be localized to the nucleus with a certainty of 0.9800.

Table 32B. Encoded NOV32 protein sequence (SEQ ID NO:132)

MATGADVDRDILELGGPEGDAASGTISKDI INPDKKSKKSSETLTFRPEGMHREVYALLYSDKNKGSCLL
SRMQEDLKSFAFGHDFLAIGDAPLLPSDTGQGYRTVKAKLGSKKVRPWKMPFTNPARKDGAMFFHWRRAA
EEGKDYPFARFNKTQVPVYSEQEQYLYLHDDAWTKAETDHLFDLSRRFDLRFVVIHDDRYDHQQFKRSVED
LKERYYHICAKLANVRAPGTDLKI PVFDAGHERRRKEQLERLYNRTPEQVAEEEEYLLQELRKIEARKKERE
KRSQDLQKLITAADTTAEQRRTERKAPKKKLPQKKEAEKPAVPETAGIKFPDFKSAGVTLRSQRMKLPSSVG
QKKIKALEQMLLELGVELSPTPTEELVHMFNELRSDLVLLYELKQACANCEYELQMLRHRHEALARAGVLGG
PATPASGPGPASAEPAVTEPGLGPDPKDTI IDVVGAPLTPNSRKRRESASSSSSVKKAKKP

The disclosed NOV32 amino acid sequence has 401 of 401 amino acid residues (100%) identical to, and 401 of 401 amino acid residues (100%) similar to, the 467 amino acid residue ptnr:SPTREMBL-ACC:Q9NPF5 protein from *Homo sapiens* (Human) (Hypothetical 53.0 Kda Protein (Dnmt1 Associated Protein-1) ($E = 1.3e^{-248}$).

NOV32 is predicted to be expressed in at least Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV32 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 32C.

Table 32C. BLAST results for NOV32

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7243231 dbj BAA92663.1 (AB037846)	KIAA1425 protein [<i>Homo sapiens</i>]	495	446/473 (94%)	446/473 (94%)	0.0
gi 13123776 ref NP_061973.1 (NM_019100)	DNA methyltransferase 1-associated protein 1 [<i>Homo sapiens</i>]	467	446/473 (94%)	446/473 (94%)	0.0
gi 12052838 emb CAB66592.1 (AL136657)	hypothetical protein [<i>Homo sapiens</i>]	467	443/473 (93%)	445/473 (93%)	0.0
gi 12963557 ref NP_075667.1 (NM_023178)	DNMT1 associated protein-1 [<i>Mus musculus</i>]	468	437/474 (92%)	438/474 (92%)	0.0
gi 12805675 gb AAH02321.1 AAH02321 (BC002321)	Unknown (protein for IMAGE:3594236) [<i>Mus musculus</i>]	451	420/457 (91%)	421/457 (91%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 32D.

(The following information was obtained from the records of the Federal Bureau of Investigation, Department of Justice.)

277

			370	380	390	400	410	420	
	NOV32	333	392
5	gi 7243231	335	394
	gi 13123776	307	366
	gi 12052838	307	366
	gi 12963557	307	366
	gi 12805675	290	349
10			430	440	450	460	470	480	
	NOV32	393	452
	gi 7243231	395	454
15	gi 13123776	367	426
	gi 12052838	367	426
	gi 12963557	367	426
	gi 12805675	350	409
20			490	500	510	520			
	NOV32	453	493
	gi 7243231	455	495
	gi 13123776	427	467
	gi 12052838	427	467
25	gi 12963557	427	468
	gi 12805675	410	451

Methylation of CpG islands is associated with transcriptional silencing and the formation of nuclease-resistant chromatin structures enriched in hypoacetylated histones. Methyl-CpG-binding proteins, such as MeCP2, provide a link between methylated DNA and hypoacetylated histones by recruiting histone deacetylase, but the mechanisms establishing the methylation patterns themselves are unknown. Whether DNA methylation is always causal for the assembly of repressive chromatin or whether features of transcriptionally silent chromatin might target methyltransferase remains unresolved. Mammalian DNA methyltransferases (DNMT) show little sequence specificity *in vitro*, yet methylation can be targeted *in vivo* within chromosomes to repetitive elements, centromeres and imprinted loci. This targeting is frequently disrupted in tumour cells, resulting in the improper silencing of tumour-suppressor genes associated with CpG islands. Robertson et al. (Nat Genet 2000, 25:338-42) have shown that the predominant mammalian DNA methyltransferase, DNMT1, co-purifies with the retinoblastoma (Rb) tumour suppressor gene product, E2F1, and HDAC1 and that DNMT1 cooperates with Rb to repress transcription from promoters containing E2F-binding sites. These results establish a link between DNA methylation, histone deacetylase and sequence-specific DNA binding activity, as well as a growth-regulatory pathway that is disrupted in nearly all cancer cells. Recently, Rountree et al. (Nat Genet, 2000, 25:269-77) have shown that the non-catalytic amino terminus of DNMT1 binds to HDAC2 and a new protein, DMAP1 (for DNMT1 associated protein), and can mediate transcriptional repression. DMAP1 has intrinsic transcription repressive activity, and binds to the transcriptional co-repressor TSG101.

DNMT1 is targeted to replication foci through interaction with the far N terminus of DNMT1 throughout S phase, whereas HDAC2 joins DNMT1 and DNMT1 only during late S phase, providing a platform for how histones may become deacetylated in heterochromatin following replication. Thus, DNMT1 not only maintains DNA methylation, but also may directly target, in a heritable manner, transcriptionally repressive chromatin to the genome during DNA replication.

The NOV32 nucleic acid of the invention encoding a DNMT1 associated protein-1 (DNMT1)-like protein includes the nucleic acid whose sequence is provided in Table 32A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 32A while still encoding a protein that maintains its DNMT1 associated protein-1 (DNMT1)-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 3% of the residues may be so changed.

The NOV32 protein of the invention includes the DNMT1 associated protein-1 (DNMT1)-like protein whose sequence is provided in Table 32B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 32B while still encoding a protein that maintains its DNMT1 associated protein-1 (DNMT1)-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 9% of the bases may be so changed.

The NOV32 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: cancers such as breast cancer, colorectal cancers, lung cancer, liver cancer, pancreatic cancer, prostate cancer, stomach cancers,

developmental syndromes, Fragile X and Rett and other diseases, disorders and conditions of the like.

NOV32 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. For example the disclosed NOV32 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV33

A disclosed NOV33 nucleic acid of 7693 nucleotides (also referred to as CG56688-01) encoding a novel Notch1-like protein is shown in Table 33A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TAA codon at nucleotides 7669-7671. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 33A, and the start and stop codons are in bold letters.

Table 33A. NOV33 Nucleotide Sequence (SEQ ID NO:133)

ATGCCGCCGCTCCTGGCGCCCTGCTCTGCCTGGCGCTGCTGCCCGCTCGCCGCACGAGGCCCGCGATG
CTCCAGCCCGGTGAGACCTGCCTGAATGGCGGGAAGTGTGAAGCGGCCAATGGCAGGAGCGCTGCGTCT
GTGGCGGGCCTTCGTGGGCCCGCGATGCCAGGACCCCAACCCGTGCCTCAGCACCCCTGCAAGAACGCC
GGGACATGCCACGTGGTGGACCGCAGAGGCGTGGCAGACTATGCCTGCAGCTGTGCCCTGGGCTTCTCTGG
GCCCCCTCTGCCTGACACCCCTGGACAACGCTGCCTCACCAACCCCTGCCGCAACGGGGGCACCTGCGACC
TGCTCACGCTGACGGAGTACAAGTGCCGCTGCCCGCCCGGCTGGTCAGGGAAATCGTGCCAGCAGGCTGAC
CCGTGCGCCTCCAACCCCTGCGCCAACGGTGGCCAGTGCCTGCCCTTCGAGGCCCTCCTACATCTGCCACTG
CCCCACCCAGCTTCCATGGCCCCACCTGCCGGCAGGATGTCAACGAGTGTGGCCAGAAGCCCCGGGCTTTGCC
GCCACGGAGGCACCTGCCACAACGAGGTGCGCTCCTACCGCTGCGTCTGCCGCGCCACCCACACTGGCCCC
AACTGCGAGCGGCCCTACGTGCCCTGCAGCCCTCGCCCTGCCAGAACGGGGGCACCTGCCGCCCCACGGG
CGACGTCACCCACGAGTGTGCCTGCCTGCCAGGCTTCACGGCCAGAAGTGTGAGGAAAATATCGACGATT
GTCCAGGAAACAAGTGAAGAACGGGGGTGCCTGTGTGGACGGCGTGAACACCTACAAGTCCCGTGCCCG
CCAGAGTGGACAGGTCACTACTGTACCGAGGATGTGGACGAGTGCAGCTGATGCCAAATGCTGCCAGAA
CGGCGGGACCTGCCACAACACCCACGGTGGCTACAAGTGCCTGTGTCAACGGCTGGACTGGTGGAGT
GCAGCGAGAACATTGATGACTGTGCCAGCGCGCCTGCTTCCACGGCGCCACCTGCCATGACCGTGTGGCC
TCTTTTACTGCGAGTGTCCCATGGCCGCACAGGTCTGCTGTGCCACCTCAACGACGCATGCATCAGCAA
CCCCTGTAACGAGGGCTCCAAGTGCAGACCAACCCCTGTCAATGGCAAGGCCATCTGCACCTGCCCTCGG
GGTACAGGGCCCGGCTGCAGCCAGGACGTGGATGAGTGTGCTCGCTGGGTGCCAACCCCTGCCGAGCATGCG
GGCAAGTGCATCAACAGCTGGGCTCCTTCGAGTGCCAGTGTCTGCAGGGCTACACGGGCCCCCGATGCCGA
GATCGACGTCAAAGTGCCTCTCGAACCCGTGCCAGAACGACGCCACCTGCCTGGACAGATTGGGGAGT
TCCAGTGCATGTGCATGCCCGGCTACGAGGTGTGCACTGCGAGGTCAACACAGACGAGTGTGCCAGCAGC
CCCTGCTGCACAATGGCCGCTGCCTGGACAAGATCAATGAGTTCCAGTGCAGTGCCTGCCAGGCTTAC
TGGGCATCTGTGCCAGTACGATGTGGACGAGTGTGCCAGCACCCCTGCAAGAATGGTGCCAAGTGCCTGG
ACGGACCCCAACACTTACACCTGTGTGTGCAGGAAGGTACACGGGACGCACTGCGAGGTGGACATCGAT
GAGTGCAGCCCCGACCCCTGCCACTACGGCTCCTGCAAGGACGGCGTGCACACCTTACCTGCGCTTGCCTG
CCAGGCTACACGGGGCCACCACTGCGAGACCAACATCAACGAGTGTCCAGCCAGCCCTGCCGCCACGGGG
GCACCTGCCAGGACCGCGACAACGCTACCTCTGCTTCTGCCTGAAGGGGACCACAGGACCCCACTGCGAG
ATCAACCTGGATGACTGTGCCAGCAGCCCCCTGCGACTCGGGCACCTGTCTGGACAAGATCGATGGCTACGA

GTGTGCTGTGAGCCGGGCTACACAGGGAGCATGTGTAACATCAACATCGATGAGTGTGCGGGCAACCCCT
 GCCACAACGGGGGACCTGCGAGGACGGCATCAATGGCTTACCTGCCGCTGCCCGAGGGCTACACGAC
 CCCACCTGCCTGTCTGAGGTCAATGAGTGCACACGCAACCCCTGCGTCCACGGGGCTGCCGGGACAGCCT
 CAACGGGTACAAGTGCAGCTGTGACCTGGGTGGAGTGGGACCAACTGTGACATCAACAACAACGAGTGTG
 AATCCAACCCCTGTGTCAACGGCGGCACCTGCAAGACATGACCAGTGCTACGTGTGCACCTGCCGGGAG
 GGCTTACGGCTCCCAACTGCCAGACCAACATCAACGAGTGTGCGTCCAACCCATGTCTGAACAAGGGCAC
 GTGTATTGACAGCTTGGCGGGTACAAGTGCAACTGCCCTGCTGCCCTACACAGGTGCCACGTGTGAGGTGG
 TGCTGGCCCCGTGTGCCCCAGCCCCCTGCAGAAACGGCGGGGAGTGCAGGCAATCCGAGGACTATGAGAGC
 TTCTCCTGTGTCTGCCCCAGGGCTGGGGCCAAAGGGCAGACCTGTGAGGTGACATCAACGAGTGCCTTCT
 GAGCCCGTGGCGGACGGCGCATCCTGCCAGAACCCACGGCGGCTACCGCTGCCACTGCCAGGCGGGCT
 ACAGTGGGCGCAACTGCGAGACCGACATCGAGCTGCCGGCCCAACCCGTGTGCACACGGGGGCTCCCTGC
 ACAGACGGCATCAACACGGCCTTCTGCGACTGCTGCCCCGCTTCCGGGGCACTTCTGTGAGGAGGACAT
 CAACGAGTGTGCCAGTGACCCCTGCCGCAACGGGGCCAACTGCACGGACTGCGTGGACAGCTACACGTGCA
 CCTGCCCGCAGGCTTACGCGGGATCCACTGTGAGAACAACACGCCCTGACTGCACAGAGAGCTCCTGCTT
 AACGGTGGCACCTGCGTGGACGGCATCAACTCGTTACCTGCCTGTGTCCACCCGGCTTACGGGCAGCTA
 CTGCCAGCACGATGTCAATGAGTGCAGCTCACAGCCCTGCCTGCATGGCGGCACCTGTGAGGACGGCTGCG
 GCTCCTACAGGTGCACCTGCCCGCAGGGCTACACTGGCCCCAACTGCCAGAACCTTGTGACTGGTGTGAC
 TCCCTGCCCTGCAAGAACGGCGGCAATGCTGGCAGACCCACACCCAGTACCGTGGAGTGGCCAGCGG
 CTGGACCGGCTTACTGCGAGCTGCCAGCGTGTCTGTGAGGTGGTGGCAGCGACAAGGTGTGACG
 TTGCCCCCTGTGCCAGCATGGAGGGCTGTGTGTGGACGGGGCAACCGCACCCTGCCGCTGCCAGGGC
 GGCTACACAGGAGCTACTGTGAGGACCTGGTGGACGAGTGTCTACCCAGCCCCCTGCCAGAACGGGGCCAC
 CTGACGGACTACTTGGGCGGCTACTCTGCAAGTGCCTGGTGGCGGGCTACCAAGGGGTGAATGCTGAGG
 AGATCGACGAGTGCCTCTCCACCCCTGCCAGAACGGGGGCACTGCCTCGACCTCCCCAACACCTACAAG
 TGCTCCTGCCACGGGGCACTCAGGGTGTGCACTGTGAGATCAACGTGGACGACTGCAATCCCCCGCTTGA
 CCGCGTGTCCGGAGCCCAAGTGCTTTAACAACGGGCACTGCGTGGACAGGTGGGCGGCTACAGTGTGA
 CCTGCCCGCGGGCTTCTGGGTGAGCGCTGTGAGGGGGATGTCAACGAGTGCCTGTCCAATCCTGCGAC
 GCCCGTGGCACCCAGAACTGCGTGCAGCGCGTCAATGACTTCCACTGCGAGTGGCGTGTGTGTACACCGG
 GCGCGCTGCGAGTCCGTCAATGGCTGCAAGGCAAGCCCTGCAAGAATGGGGGCACTTGCCTGGTGG
 CCGCAACACCGCCCGCGGGTTTACTGTGCAAGTGCCTGCGGGCTTCGAGGGCGCGCAGTGTGAGATGAC
 GCTCGTACCTGCGGCAGCCTGCGTGCCTCAACGGCGGCACATGCATCTCCGGCCCGCGCAGCCCCACCTG
 CCTGTGCTTGGGCCCCCTTACGGGCCCCGAATGCCAGTTCCCGGCCAGCAGCCCCCTGCCGTGGCGGCAACC
 CTGCTACAACAGGGGACCTGTGAGCCACATCCGAGAGCCCCCTTCTACCGTTGCCTGTGCCCCGCCAAA
 TTAACCGGGCTTGTGTGCACATCCTGGACTACAGCTTCCGGGGTGGGGCGGGCGGCGACCTCCCGCC
 GCTGATCGAGGAGCGTGCAGCTGCCCGAGTGCAGGAGGACGGGGCAACAAGGTCTGCAGCCTGCAGT
 GCAACAACACCGCTGCGGCTGGGACGGCGGTGACTGCTCCCTCAACTTCAATGACCCCTGGAGAAGTGC
 ACGCAGTCTCTGCACTGCTGGAAGTACTTCACTGACGGCCACTGTGACAGCCAGTGAACCTAGCCGGCTG
 CCTCTCGACGGCTTTGACTGCCAGCTGCGGAAGGCCAGTGCAACCCCTGTACAGCAAGTATGAGAGG
 ACCACTTCAGCGACGGGCACTGCGACAGGGGTGCAACAGCGCGGAGTGCAGTGGGACGGGCTGGACTGT
 GCGGAGCATGTACCCGAGAGGCTGGCGGCGGCGACGCTGGTGGTGGTGGTGTGATGCCGCGGAGCAGCT
 GCGCAACAGCTCCTTCCACTTCTGCGGGAGCTCAGCCGCTGCTGCACACCAACAGTGTCTTCAAGCTG
 ACGCACACGGCCAGCAGATGATCTTCCCCCTACTACGGCCCGGAGGAGGAGCTGCGCAAGCACCCCATCAAG
 CGTGCCCGCGAGGGCTGGGCGGCACCTGACGCCCTGCTGGGCCAGGTGAAGGCCTCGCTGCTCCTGGTGG
 CAGCGAGGGTGGGCGGCGGCGGAGGAGCTGGACCCATGGACGTCCGCGGCTCCATCTGCTTACCTGGAGA
 TTGACAACCGGCAGTGTGTGACGGCCTCCTCGAGTGTCTCCAGAGTGCCACCGCATGTGCGCGCATCTG
 GGAGCGCTCGCTCGTGGGCGAGCCTCAACATCCCCCTACAAGATCGAGGCCGTGCAGAGTGAGACCGTGA
 GCCGCCCCCGCGCGCGCAGCTGCACTTCAATGTAGTGGCGGCGGCGGCTTGTGCTTCTGTTCTTCTG
 GCTGCGGGGTGCTGCTGTCCCGCAAGCGCGGCGGCGGAGCATGGCCAGCTCTGGTTCCCTGAGGGCTTCAA
 GTGTCTGAGGCCAGCAAGAAGAAGCGGGGAGCCCCCTCGCGGAGGACTCCGTGGGCTCAAGCCCTGA
 GAACGCTTCAGACGGTGCCTCATGGACGACAACAGAATGAGTGGGGGACGAGGACCTGGAGACCAAGA
 AGTTCGGTTCGAGGAGCCCGTGGTTCGCTGACCTGGACGACGACAGACACCGGCGAGTGGACTCAG
 CAGCACCTGGATGCCGCTGACCTGCGCATGTCTGCCATGGCCCCCACACCGCCAGGTTGAGGTGACGC
 CGACTGCATGGACGTCAATGTCCGCGGGCTGATGGCTTACCCCGCTCATGATCGCCTCCTGACGCGGGG
 GCGGCTGGAGACGGGCAACAGCGAGGAAGAGGAGGACGCGCGGCGCTCATCTCGACTTCATCTACAG
 GCGCCAGCCTGCACAACAGACAGACCGCACGGGCGAGACCGCTTGACCTGGCCGCGGCTACTACAG
 CTCTGATGCCGCCAAGCGCCTGCTGGAGGCCAGCGCAGATGCCAACATCCAGGACAACATGGGCGGACCC
 CGCTGCATGCGGCTGTGTCTGCCGACGCACAAGGTGTCTTCCAGATCCTGATCCGGAACAGGGCCACAGAC
 CTGGATGCCCGCATGCATGATGGCACAACCTCACTGATCCTGGCTGCCCGCTGGCGTGGAGGGCATGCT
 GGAGGACCTCATCAACTCACACGCGGACGTCAACGCGGTAGATGACCTGGGCAAGTCCGCCCCCTGCACTGGG
 CCGCGCGGTGAACAATGTGGATGCCGAGTTGTGCTCTTGAAGAACGGGGCTAACAAAGATATGAGAAC
 AACAGGGAGGAGACACCCCTGTTTCTGGCCGCGCGGAGGGGAGCTACGAGACCGCAAGGTGTGTGGA
 CCACTTTGCCAACCGGGACATCACGGATCATATGGACCCCTGCCGCGGACATCGCACAGGAGCGCATGC
 ATCAGGACATCGTGAAGCTGCTGGACGAGTACAACCTGGTGGCAGCCCGCAGCTGCACGGAGCCCCGCTG
 GGCGGACGCCCCACCCCTGTGCCCCCGCTGCTGCCCAACGGCTACCTGGGAGCCTCAAGCCCGGCT
 GCAGGGCAAGAAGTCCGCAAGCCAGCAGCAAGGCCCTGGCTGTGGAAGCAAGGAGGCCAAGGACCTCA
 AGGCACGGAGGAAGAAGTCCAGGATGGCAAGGGCTGCCTGCTGGACAGCTCCGCGCATGCTCTGCCCCGTG
 GACTCCTGGAGTCAACCCATGGCTACCTGTGACAGTGGCTCGCCGCCACTGCTGCCCTCCCCGTTCCA
 GAGTCTCCGTCCGTGCCCTCAACCACTGTGGGATGCCCGACACCCACCTGGGCACTGGGCACTGA
 ACGTGGCGGCAAGCCCGAGATGGCGGCGTGGTGGGGGCGGCGGCTGGCCTTGTGACTGGCCACCT
 CGTCTCTCCACCTGCCTGTGGCCTTGTGCCACAGCACCGTCTGGACTCAGCAGCGGAGGGGCTGAA

TTTCACTGTGGGCGGGTCCACCAGTTGAATGGTCAATGCGAGTGGCTGTCCCGGCTGCAGAGCGGCATGG
 TGCCGAACCAATACAACCTCTGCGGGGGAGTGTGGCACCAGGCCCCCTGAGCACACAGGCCCCCTCCCTG
 CAGCATGGCATGGTAGGCCCGCTGCACAGTAGCCTTGCTGCCAGCGCCCTGTCCAGATGATGAGCTACCA
 GGGCTGCCAGCACCAGGCTGGCCACCCAGCCTCACCTGGTGCAGACCCAGCAGGTGCAGCCACAAAACCT
 TACAGATGCAGCAGCAGAACCTGCAGCCAGCAACATCCAGCAGCAGCAAGCCTGCAGCCGCCACCACCA
 CCACCACAGCCGACCTTGGCGTGAGCTCAGCAGCCAGCGGCCACCTGGGCGGAGCTTCTGAGTGGAGA
 GCGGAGCCAGGCAGACGTGCAGCCACTGGGCCCCAGCAGCCTGGCGGTGCACACTATTCTGCCCCAGGAGA
 GCGCCGCCCTGCCACGTGCTGCCATCCTCGCTGGTCCCACCCGTGACCGCAGCCAGTTCCTGACGCCC
 CCTCGCAGCACAGCTACTCCTCGCTGTGGACAACCCCCAGCCACCAGCTACAGGTGCTGAGCAGCC
 CTTCTGACCCCTTCGCGGAGTGCGCCGACCAATGGTCGTCTCTGTCGCGGCACTCTAATGTGTCTGACT
 GGTCTGAGGGCGTGTCTGCGCCCCGACCTCCATGCAGTCCAGATCGCGCGCATCCCGAGGCGTTCAAG
TAATAGCTCGAGGTGCCAGCAGTC

The NOV33 nucleic acid has 7670 of 7693 bases (99%) identical to a gb:GENBANK-
 ID:AF308602|acc:AF308602.1 mRNA from *Homo sapiens* (NOTCH 1 (N1) mRNA, complete
 cds) (E = 0.0).

5 A disclosed NOV33 polypeptide (SEQ ID NO:134) encoded by SEQ ID NO:133 is
 2556 amino acid residues and is presented using the one-letter code in Table 33B. Signal P,
 Psort and/or Hydropathy results predict that NOV33 contains a signal peptide and is likely to
 be localized to the plasma membrane with a certainty of 0.4600. The most likely cleavage site
 for a NOV33 peptide is between amino acids 18 and 19: ALA-AR.

Table 33B. Encoded NOV33 protein sequence (SEQ ID NO:134)

MPPLLAPLLCLALLPALAARGPRCSQPGETCLNGGKCEAANGTEACVCGGAFVGPQCQDPNPCLSTPCKNAG
 TCHVVDRRGVADYACSCALGFSGLCLTPLDNACLTNPCRNGGTCDDLTLTEYKRCRCPGWSGKSCQADPC
 ASNPCANGGQCLPFEASYICHPPSFHGPTCRQDVNECGQKPGLCRHGGTCHNEVGSYRCVCRATHTGPNC
 RPYVPCSPSPCQNGGTCTGDTHECACLPFTGQNCENIDDCPGNNCKNGGACVDGVTNCPPEWT
 GQYCTEDVDEQCLMPNACQNGGTCHNTHGGYNCVNVNGWTGEDCSENIDDCASAAFCFHGATCHDRVASFYCE
 CPHGRTGLLCHLNDACISNPCNEGSNCDTNPVNGKAICTCPSGYTGPAQSQDVDECSLGNPCEHAGKCI
 LGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCPGYEGVHCEVNTDECASSPCLHNGR
 CLDKINEFQCECPTGFTGHLQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCH
 YGSKDGVATFTCLCRPGYTGHHCEETNINECSSQPCRHHGTCDQDRDNAYLCFLKGTTPNCEINLDDCASS
 PCDSGTCLDKIDGYEACCEPGYTGSMCNINIDECAGNPCHNGGTCEGTINGFTCRCEGYHPTCLSEVNEC
 NSNPCVHGACRDSLNGYKDCDPCGWSGTNCDINNNECESNPCVNGGTCKDMTSGYVCTCREGFSGPNCQNTI
 NECASNPCNLKGTCTIDVAGYKCNCLLPYTATCEVVLAPCAPSPCRNGGECRQSEDEYESFCVCPTAGAKG
 QTCEVDINECVLSPCRHGASCQNTHGGYRCHCQAGYSGRNCETDIDDCRPNPCHNGGSCSTDGINTAFCDCLP
 GFRGTFCCEEDINECASDPCRNGANCTDCVDSYTCPCAGFSGIHCENNTPDCTESSCFNGGTCTVDGINSFTC
 LCPPGFTGSYQCHDVNECDSQPCLHGGTCQDGCYSYRCTCPQGYTGPNQNLVHWCDSPPCKNGGKQWQHT
 QYRCECPSGWTGLYCDVPSVSCEVAAQRQGVVARLCQHGGCLVDAGNTHHCRQAGYTGSYCEDLVDECSP
 SPCQNGATCTDYLGGYSCKCVAGYHGVNCSEIDECLSHPCQNGGTCLDLNPTYKCSCPRGTQGVHCEINVD
 DCNPPVDVPSRSPKCFNNGTCTVDQVGGYSCTCPPGFVGRCEDVNECLSNPCDARGTQNCVQRVNDFHCEC
 RAGHTGRRCESVINGCKGKPCKNGGTCAVASNTARFICKCPAGFEGATCENDARTCGSLRCLNNGGTCTISGP
 RSPTCLCLGPFTGPECQFPASSPCLGPNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFSGGAGRD
 IPPPLIEEACELPECQEDAGNKVCSLQCNHACGWDGDCSLNFNDPWKNCTQSLQCWKYFSDGHCDSDQCN
 AGCLDFGDFCQRAEGQCNPPLYDQYCKDHFSHGCDQGCNSAECEWDGLDCAEHVPERLAAGTLVVVLMPE
 QLRNSSFFHLRELRLVHTNVVFKRDAHQQMI FPIYGGREEELRKHPIKRAEGWAAPDALLGQVKASLLPG
 GSEGGRRRLDPMQVRSIVYLEIDNRQCVQASSQCFQSATDVAAFLGALASLGLSNIPIYKIEAVQSETVE
 NHPGLMPDTHLIGIHLNVAAPKPEMAALGGGRLAFETGTPRLSHLPVASTSTVLGSSSGALNFTVGGSTS
 ASDGALMDDNQNEWGDELETKKFRFEFPVLPDLDDQTDHRQWTQQLDAADLRMSAMAPTTPQGEVDADC
 MDVNVVRGPDGFTPLMIASCSGGLETGNSEEEEDAPAVISDFIYQGASLHNQTDRTGETALHLAARYSRSDA
 AKRLLASADANIQDNMGRTPHLAAVSADAQGVFQILIRNRATDLDARMHDGTTPLILAAARLAVEGMLEDLI
 NSHADVNAVDDLKKSALHWAAAVNNVDAAVLLKNGANKDMQNNREETPLFLAAREGSYETAKVLLDHFANR
 DITDHMDRLPRDIAQERMHHDIIVRLDEYNLVRSPQLHGAPLGGTPTLSPPLCSPNGYLSLKPVGQKKVR
 KPSSKGLACGSKEAKDLKARRKKSQDGKGLDSSGMLSPVDSLESPLHGYLSDVASPPLPSFQQSPSPVPL
 NHLPGMPDTHLIGIHLNVAAPKPEMAALGGGRLAFETGTPRLSHLPVASTSTVLGSSSGALNFTVGGSTS
 LNGQCEWLSRLQSGMVPNQYNPLRGSVAPGPLSTQAPSLQHGMMVGLHSSLAASALSQMSYQGLPSTRLAT
 QPHLVQTQVQVQPNLQMQQNLQPANIQQQQSLQPPPPPPQPHLVGSSAASGHLGRSFLSGEPSQADVQPLG
 PSSLAHVHTILPQESPALEPSTLSPSSLVPPVTAQFLTPPSQHSYSSPVDNTPSHQLQVPEHFFLTPSPESPQ

WSSSSPHSNVSDWSEGVSSPPTSMQSQIARIPFAFK

The disclosed NOV33 amino acid sequence has 2543 of 2556 amino acid residues (99%) identical to, and 2545 of 2556 amino acid residues (99%) similar to, the 2556 amino acid residue ptnr:TREMBLNEW-ACC:AAG33848 protein from *Homo sapiens* (Human) (Notch 1) (E = 0.0).

NOV33 is predicted to be expressed in at least Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV33 is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF308602|acc:AF308602.1) a closely related *Homo sapiens* NOTCH 1 (N1) mRNA, complete cds homolog in species *Homo sapiens*: brain.

NOV33 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 33C.

Table 33C. BLAST results for NOV33

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 11275980 gb AAG33848.1 AF308602_1 (AF308602)	NOTCH 1 [<i>Homo sapiens</i>]	2556	2543/2556 (99%)	2545/2556 (99%)	0.0
gi 107215 pir A40043	notch protein homolog TAN-1 precursor - human	2555	2537/2556 (99%)	2541/2556 (99%)	0.0
gi 1171777 sp P46531 NTC1_HUMAN	NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLOCATION-ASSOCIATED NOTCH PROTEIN TAN-1)	2444	2429/2444 (99%)	2431/2444	0.0
gi 6093542 sp Q07008 NTC1_RAT	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR	2531	2301/2557 (89%)	2401/2557 (92%)	0.0
gi 112074 pir S18188	notch protein homolog - rat	2531	2300/2557 (89%)	2399/2557 (92%)	0.0

[illegible]284

	gi 6093542	301	LMPNACQNAAGTCHNSHGGSYNVCVNGWTGEDCS	360
	gi 112074	301	LMPNACQNAAGTCHNSHGGSYNVCVNGWTGEDCS	360
5			370 380 390 400 410 420	
	NOV33	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
	gi 11275980	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
	gi 107215	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
10	gi 1171777	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
	gi 6093542	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
	gi 112074	361	CPHGRTGLLCHLNDAACISNPNCEGSNC	420
15			430 440 450 460 470 480	
	NOV33	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
	gi 11275980	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
	gi 107215	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
	gi 1171777	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
20	gi 6093542	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
	gi 112074	421	NPCEHAGKCIINTLGSFECQCLQGYTGPRCEIDVNECVSNPCQNDATCLDQIGEFQCMCMP	480
25			490 500 510 520 530 540	
	NOV33	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	540
	gi 11275980	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	540
	gi 107215	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	539
	gi 1171777	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	540
	gi 6093542	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	540
30	gi 112074	481	GYEGVHCEVNTDECASSPCLHNGRCLDKINEFQCECPTGFTGHLCOYDVDECASTPCKNG	540
35			550 560 570 580 590 600	
	NOV33	541	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	600
	gi 11275980	541	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	600
	gi 107215	540	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	599
	gi 1171777	541	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	600
	gi 6093542	541	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	600
	gi 112074	541	AKCLDGPNTYTCVCTEGYTGTHCEVDIDECDPDPCHYGSCKDGVATFTCLCRPGYTGHHC	600
40			610 620 630 640 650 660	
	NOV33	601	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	660
	gi 11275980	601	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	660
	gi 107215	600	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	659
45	gi 1171777	601	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	660
	gi 6093542	601	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	660
	gi 112074	601	ETNINECSSQPCLRHGCTCQDRDNYLFCFLKGTGTPNCEINLDDCASSPCDSGTCLDKID	660
50			670 680 690 700 710 720	
	NOV33	661	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	720
	gi 11275980	661	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	720
	gi 107215	660	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	719
	gi 1171777	661	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	720
55	gi 6093542	661	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	720
	gi 112074	661	GYECACEPGYTGSNCNINIDECAGNPCHNGGTCE	720
60			730 740 750 760 770 780	
	NOV33	721	NSNPCVHGACRDSLNGYKCDP	780
	gi 11275980	721	NSNPCVHGACRDSLNGYKCDP	780
	gi 107215	720	NSNPCVHGACRDSLNGYKCDP	779
	gi 1171777	721	NSNPCVHGACRDSLNGYKCDP	780
	gi 6093542	721	NSNPCVHGACRDSLNGYKCDP	780
65	gi 112074	721	NSNPCVHGACRDSLNGYKCDP	780
70			790 800 810 820 830 840	
	NOV33	781	EGFSGPNQCOTNINECASNPCLNKGTCID	840
	gi 11275980	781	EGFSGPNQCOTNINECASNPCLNKGTCID	840

5	gi 107215	780	EGFSGPNCQTNINECASNPCLNKGTCTIDDVAGYKCNCLLPYTGATCEVVLAPCAPSPCRN	839
	gi 1171777	781	EGFSGPNCQTNINECASNPCLNKGTCTIDDVAGYKCNCLLPYTGATCEVVLAPCAPSPCRN	840
	gi 6093542	781	EGFSGPNCQTNINECASNPCLNKGTCTIDDVAGYKCNCLLPYTGATCEVVLAPCAPSPCRN	840
	gi 112074	781	EGFSGPNCQTNINECASNPCLNKGTCTIDDVAGYKCNCLLPYTGATCEVVLAPCAPSPCRN	840
10	NOV33	841	GGECROSEDIYESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	900
	gi 11275980	841	GGECROSEDIYESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	900
	gi 107215	840	GGECROSEDIYESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	899
	gi 1171777	841	GGECROSEDIYESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	900
	gi 6093542	841	SGVCKESEDYIESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	899
	gi 112074	841	SGVCKESEDYIESFSCVCPYTAGAKGQTCEVDINECVLSPCRHGASCONTHGXYRCHCQAGY	899
20	NOV33	901	SGRNCETDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	960
	gi 11275980	901	SGRNCETDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	960
	gi 107215	900	SGRNCETDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	959
	gi 1171777	901	SGRNCETDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	960
	gi 6093542	900	TGRNCESDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	959
	gi 112074	900	TGRNCESDIDDCRPNPCHNGGSCDTGINTAFCDCLPGFRGTFCCEEDINECASDPCRNGAN	959
30	NOV33	961	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1020
	gi 11275980	961	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1020
	gi 107215	960	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1019
	gi 1171777	961	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1020
	gi 6093542	960	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1019
	gi 112074	960	CTDCVDSYTCCTCPAGFSGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCO	1019
40	NOV33	1021	HVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1080
	gi 11275980	1021	HVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1080
	gi 107215	1020	HVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1079
	gi 1171777	1021	HVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1080
	gi 6093542	1020	YDVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1079
	gi 112074	1020	YDVNECDSPCLHGGTCQDGRGLHRTCPQGYTGPNQCNLVHWCDSPPCKNGGKCQWQHT	1079
55	NOV33	1081	QYRCECPSGWTGLYCDVPSVSCEVAAQROGVDVARLCOHGGLCVDAGNTHHCRCQAGYTG	1140
	gi 11275980	1081	QYRCECPSGWTGLYCDVPSVSCEVAAQROGVDVARLCOHGGLCVDAGNTHHCRCQAGYTG	1140
	gi 107215	1080	QYRCECPSGWTGLYCDVPSVSCEVAAQROGVDVARLCOHGGLCVDAGNTHHCRCQAGYTG	1139
	gi 1171777	1081	QYRCECPSGWTGLYCDVPSVSCEVAAQROGVDVARLCOHGGLCVDAGNTHHCRCQAGYTG	1140
	gi 6093542	1080	QYHCECRSGWTGFNCVDVLSVSCEVAAQKRGIDVTLLCOHGGLCVDEEDKHCHCQAGYTG	1139
	gi 112074	1080	QYHCECRSGWTGFNCVDVLSVSCEVAAQKRGIDVTLLCOHGGLCVDEEDKHCHCQAGYTG	1139
65	NOV33	1141	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1200
	gi 11275980	1141	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1200
	gi 107215	1140	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1199
	gi 1171777	1141	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1200
	gi 6093542	1140	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1199
	gi 112074	1140	SYCEDLVDECSPSPCQNGATCTDYLGGYSCKCVAGYHGVCNCEEIDECLSHPCQNGGTCL	1199
70	NOV33	1201	DLPNYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1260
	gi 11275980	1201	DLPNYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1260
	gi 107215	1200	DLPNYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1259
	gi 1171777	1201	DLPNYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1260
	gi 6093542	1200	DLNTYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1259
	gi 112074	1200	DLNTYKCSKSPRGTQGVHCEINVDCCNPPVDPVSRSPKCFNNGTCVDQVGGYSCTCPPGF	1259

5	NOV33	1261	VGERCEGDVNECLSNPCDARGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1320
	gi 11275980	1261	VGERCEGDVNECLSNPCDARGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1320
	gi 107215	1260	VGERCEGDVNECLSNPCDARGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1319
	gi 1171777	1261	VGERCEGDVNECLSNPCDARGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1320
	gi 6093542	1260	VGERCEGDVNECLSNPCDPRGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1319
	gi 112074	1260	VGERCEGDVNECLSNPCDPRGTQNCVQVRVNDHFCECRAGHTGRRCESVINGCKGKPKCKNG	1319
10	NOV33	13211330.....1340.....1350.....1360.....1370.....1380..... GTCAVASNTARGFICKCPAGFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGPFT	1380
	gi 11275980	1321	GTCAVASNTARGFICKCPAGFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGPFT	1380
	gi 107215	1320	GTCAVASNTARGFICKCPAGFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGPFT	1379
	gi 1171777	1321	GTCAVASNTARGFICKCPAGFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGPFT	1380
	gi 6093542	1320	GTCAVASNTARGFICKCPARFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGSFT	1379
15	gi 112074	1320	GTCAVASNTARGFICKCPARFEGATCENDARTCGSLRCLNNGGTCTISGPRSPTCLCLGSFT	1379
20	NOV33	13811390.....1400.....1410.....1420.....1430.....1440..... GPECQFPASSPCLGGNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFGGGAGRD	1440
	gi 11275980	1381	GPECQFPASSPCLGGNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFGGGAGRD	1440
	gi 107215	1380	GPECQFPASSPCLGGNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFGGGAGRD	1439
	gi 1171777	1381	GPECQFPASSPCLGGNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFGGGAGRD	1440
	gi 6093542	1380	GPECQFPASSPCVGSNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFTGAAGRD	1439
25	gi 112074	1380	GPECQFPASSPCVGSNPCYNQGTCEPTSESPFYRCLCPAKFNGLLCHILDYSFTGAAGRD	1439
30	NOV33	14411450.....1460.....1470.....1480.....1490.....1500..... IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1500
	gi 11275980	1441	IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1500
	gi 107215	1440	IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1499
	gi 1171777	1441	IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1500
	gi 6093542	1440	IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1499
	gi 112074	1440	IPPPLEIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCSLNFNDPWKNCTQSLQCWKY	1499
40	NOV33	15011510.....1520.....1530.....1540.....1550.....1560..... FSDGHCDSDQCNASAGCLFDGFDQRAEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1560
	gi 11275980	1501	FSDGHCDSDQCNASAGCLFDGFDQRAEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1560
	gi 107215	1500	FSDGHCDSDQCNASAGCLFDGFDQRAEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1559
	gi 1171777	1501	FSDGHCDSDQCNASAGCLFDGFDQRAEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1560
	gi 6093542	1500	FSDGHCDSDQCNASAGCLFDGFDCLTEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1559
	gi 112074	1500	FSDGHCDSDQCNASAGCLFDGFDCLTEGQCNPPLYDQYCKDHFSDGHCDQGCNSACEWDGL	1559
50	NOV33	15611570.....1580.....1590.....1600.....1610.....1620..... DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRELRLVLTNNVVKRDAHGGQOMIFPY	1620
	gi 11275980	1561	DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRELRLVLTNNVVKRDAHGGQOMIFPY	1620
	gi 107215	1560	DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRELRLVLTNNVVKRDAHGGQOMIFPY	1619
	gi 1171777	1561	DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRELRLVLTNNVVKRDAHGGQOMIFPY	1620
	gi 6093542	1560	DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRDVSHVLHTNNVVKRDAHGGQOMIFPY	1619
	gi 112074	1560	DCAEHVPERLAAGTLVVVVLMPPEQLRNSSFHFLRDVSHVLHTNNVVKRDAHGGQOMIFPY	1619
60	NOV33	16211630.....1640.....1650.....1660.....1670.....1680..... YGREEELRKHPIKRAAEGWAAPDALLGQVKASLLPGGSEGGRRRRELDPMDRVGSIVYLE	1680
	gi 11275980	1621	YGREEELRKHPIKRAAEGWAAPDALLGQVKASLLPGGSEGGRRRRELDPMDRVGSIVYLE	1680
	gi 107215	1620	YGREEELRKHPIKRAAEGWAAPDALLGQVKASLLPGGSEGGRRRRELDPMDRVGSIVYLE	1679
	gi 1171777	1621	YGREEELRKHPIKRAAEGWAAPDALLGQVKASLLPGGSEGGRRRRELDPMDRVGSIVYLE	1680
	gi 6093542	1620	YGREEELRKHPIKRSVVGWAT-----T--SLLPG--TNGGRORRELDPMDRVGSIVYLE	1669
	gi 112074	1620	YGREEELRKHPIKRSVVGWAT-----T--SLLPG--TNGGRHGRELDPMDRVGSIVYLE	1669
70	NOV33	16811690.....1700.....1710.....1720.....1730.....1740..... IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVQSETVEPPPPAQLHFMVY	1740
	gi 11275980	1681	IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVQSETVEPPPPAQLHFMVY	1740
	gi 107215	1680	IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVQSETVEPPPPAQLHFMVY	1739
	gi 1171777	1681	IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVQSETVEPPPPAQLHFMVY	1740
	gi 6093542	1670	IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVKSETVEPPPLPSQLHLMVY	1729
	gi 112074	1670	IDNRQCVQASSQCFQSATDVA AFLGALASLSLNIPYKIEAVKSETVEPPPLPSQLHLMVY	1729

			1750	1760	1770	1780	1790	1800	
	NOV33	1741	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1800	
5	gi 11275980	1741	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1800	
	gi 107215	1740	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1799	
	gi 1171777	1741	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1800	
	gi 6093542	1730	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1789	
	gi 112074	1730	AAAAFVLLFFVGC	VLLSRKRRRQHG	QQLWFPEGFKV	SEASKKKRREPL	GEDSVGLKPLKN	1789	
10			1810	1820	1830	1840	1850	1860	
	NOV33	1801	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1860	
	gi 11275980	1801	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1860	
	gi 107215	1800	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1859	
15	gi 1171777	1801	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1860	
	gi 6093542	1790	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1849	
	gi 112074	1790	ASDGALMDDNQNE	WGDEDLETKKFR	FEEPVLPLDDQT	DHRQWTQOHLDA	ADLRMSAMA	1849	
20			1870	1880	1890	1900	1910	1920	
	NOV33	1861	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1920	
	gi 11275980	1861	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1920	
	gi 107215	1860	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1919	
	gi 1171777	1861	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1920	
25	gi 6093542	1850	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1909	
	gi 112074	1850	PTPPQGEVDADCM	DVNVVRGPDGFT	PLMIASCSGGGLE	TGNSEEEEDAPAV	ISDFIYQGAS	1909	
30			1930	1940	1950	1960	1970	1980	
	NOV33	1921	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1980	
	gi 11275980	1921	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1980	
	gi 107215	1920	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1979	
	gi 1171777	1921	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1980	
35	gi 6093542	1910	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1969	
	gi 112074	1910	LHNQTDRTGETAL	HLAARYSRSDAAK	RLLLEASADANI	QDNMGRTPLHAAV	SADAQGVFOI	1969	
40			1990	2000	2010	2020	2030	2040	
	NOV33	1981	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2040	
	gi 11275980	1981	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2040	
	gi 107215	1980	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2039	
	gi 1171777	1981	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2040	
	gi 6093542	1970	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2029	
45	gi 112074	1970	LIRNRATDLDA	RMHDGTTPLILA	ARLAVEGMLEDL	INSHADVNAVDDL	GKSALHWAAAVN	2029	
50			2050	2060	2070	2080	2090	2100	
	NOV33	2041	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2100	
	gi 11275980	2041	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2100	
	gi 107215	2040	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2099	
	gi 1171777	2041	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2100	
	gi 6093542	2030	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2089	
	gi 112074	2030	NVDAAVVLLKNG	ANKDMQNNREET	PLFLAAREGSYET	AKVLLDHFANRDI	TDHMDRLPRD	2089	
55			2110	2120	2130	2140	2150	2160	
	NOV33	2101	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2160	
	gi 11275980	2101	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2160	
	gi 107215	2100	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2159	
60	gi 1171777	2101	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2160	
	gi 6093542	2090	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2149	
	gi 112074	2090	IAQERMHHDIVRL	LDEYNLVRSPQL	HGAPLGGTPTLS	PPPLCSPNGYLGSL	KPGVQGGKVR	2149	
65			2170	2180	2190	2200	2210	2220	
	NOV33	2161	KPSSKGLACGSKE	AKDLKARRKKSQ	DGKGCLLDSSGML	SPVDSLES	PHGYLSDVASPPLL	2220	
	gi 11275980	2161	KPSSKGLACGSKE	AKDLKARRKKSQ	DGKGCLLDSSGML	SPVDSLES	PHGYLSDVASPPLL	2220	
	gi 107215	2160	KPSSKGLACGSKE	AKDLKARRKKSQ	DGKGCLLDSSGML	SPVDSLES	PHGYLSDVASPPLL	2219	
	gi 1171777	2161	KPSSKGLACGSKE	AKDLKARRKKSQ	DGKGCLLDSSGML	SPVDSLES	PHGYLSDVASPPLL	2220	
70	gi 6093542	2150	KPSSKGLACGSKE	AKDLKARRKKSQ	DGKGCLLDSSGML	SPVDSLES	PHGYLSDVASPPLL	2209	

gi 112074		2150	KPSIKGLACSSKEAKDLKARRKKSQDGKGCLLDSSSMLSPVDSLESPhGYLSDVASPPLL	2209			
		2230	2240	2250	2260	2270	2280
5	NOV33	2221	PSPFQQSPSVPLNHLPGMPDTHLGIGHLNVAAKPEMAALGGGGRLAFETGPPRLSHLPVA	2280			
	gi 11275980	2221	PSPFQQSPSVPLNHLPGMPDTHLGIGHLNVAAKPEMAALGGGGRLAFETGPPRLSHLPVA	2280			
	gi 107215	2220	PSPFQQSPSVPLNHLPGMPDTHLGIGHLNVAAKPEMAALGGGGRLAFETGPPRLSHLPVA	2279			
	gi 1171777	2221	PSPFQQSPSVPLNHLPGMPDTHLGIGHLNVAAKPEMAALGGGGRLAFETGPPRLSHLPVA	2280			
	gi 6093542	2210	PSPFQQSPSMPLSHLPGMPDTHLGI SHLNVAAKPEMAALAGGSRLAFEP PPPRLSHLPVA	2269			
10	gi 112074	2210	PSPFQQSPSMPLSHLPGMPDTHLGI SHLNVAAKPEMAALAGGSRLAFEP PPPRLSHLPVA	2269			
		2290	2300	2310	2320	2330	2340
15	NOV33	2281	SGTSTVLGSSSSGGALNFTVGGSTSLNGQCEWLSRLQSGMVPNQYNPLRGSVAPGPLSTQA	2340			
	gi 11275980	2281	SGTSTVLGSSSSGGALNFTVGGSTSLNGQCEWLSRLQSGMVPNQYNPLRGSVAPGPLSTQA	2340			
	gi 107215	2280	SGTSTVLGSSSSGGALNFTVGGSTSLNGQCEWLSRLQSGMVPNQYNPLRGSVAPGPLSTQA	2339			
	gi 1171777	2281	SGTSTVLGSSSSGGALNFTVGGSTSLNGQCEWLSRLQSGMVPNQYNPLRGSVAPGPLSTQA	2340			
	gi 6093542	2270	SSASTVLSTNGTGAMNFTVGAPASLNGQCEWLPRLONGMVPSQYNNPLRPGVTPTGLSTQA	2329			
20	gi 112074	2270	SSASTVLSTNGTGAMNFTVGAPASLNGQCEWLPRLONGMVPSQYNNPLRPGVTPTGLSTQA	2329			
		2350	2360	2370	2380	2390	2400
25	NOV33	2341	PSLQHGMVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2400			
	gi 11275980	2341	PSLQHGMVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2400			
	gi 107215	2340	PSLQHGMVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2399			
	gi 1171777	2341	PSLQHGMVGPLHSSLAASALSQMMSYQGLPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2400			
	gi 6093542	2330	AGLQHGMVGPIHSSLSNTNLSPII-YQGLEPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2387			
	gi 112074	2330	AGLQHGMVGPIHSSLSNTNLSPII-YQGLEPSTRLATQPHLVQTQQVQPQNLMQQQNLQP	2387			
		2410	2420	2430	2440	2450	2460
30	NOV33	2401	ANIQQQQSLQPPPPPPQPHLGVSAAASCHLGRSFLSGEPSQADVQPLGPSSSLAVHTILPQ	2460			
	gi 11275980	2401	ANIQQQQSLQPPPPPPQPHLGVSAAASCHLGRSFLSGEPSQADVQPLGPSSSLAVHTILPQ	2460			
	gi 107215	2400	ANIQQQQSLQPPPPPPQPHLGVSAAASCHLGRSFLSGEPSQADVQPLGPSSSLAVHTILPQ	2459			
35	gi 1171777	2401	ANIQQQQSLQPPPPPPQPHLGVSAAASCHLGRSFLSGEPSQADV-----	2444			
	gi 6093542	2387	-----P-PSQPHLSVSSAANGHLGRSFLSGEPSQADVQPLGPSSSLPVHTILPQ	2434			
	gi 112074	2387	-----P-PSQPHLSVSSAANGHLGRSFLSGEPSQADVQPLGPSSSLPVHTILPQ	2434			
		2470	2480	2490	2500	2510	2520
40	NOV33	2461	ESPALPTSLPSSLVPPVTAQAFLTPPSQHSYSS-PVDNTPSHQLQVPEHPFLTPSPESPD	2519			
	gi 11275980	2461	ESPALPTSLPSSLVPPVTAQAFLTPPSQHSYSS-PVDNTPSHQLQVPEHPFLTPSPESPD	2519			
	gi 107215	2460	ESPALPTSLPSSLVPPVTAQAFLTPPSQHSYSS-PVENTPSHQLQVPEHPFLTPSPESPD	2518			
	gi 1171777	2444	-----	2444			
45	gi 6093542	2435	ESQALPTSLPSSMVPPMTTQFLTPPSQHSYSSSPVDNTPSHQLQVPEHPFLTPSPESPD	2494			
	gi 112074	2435	ESQALPTSLPSSMVPPMTTQFLTPPSQHSYSSSPVDNTPSHQLQVPEHPFLTPSPESPD	2494			
		2530	2540	2550			
50	NOV33	2520	QWSSSSPHSNVSDWSEGVSSPPTSMQSQTARIPEAFK	2556			
	gi 11275980	2520	QWSSSSPHSNVSDWSEGVSSPPTSMQSQTARIPEAFK	2556			
	gi 107215	2519	QWSSSSPHSNVSDWSEGVSSPPTSMQSQTARIPEAFK	2555			
	gi 1171777	2444	-----	2444			
	gi 6093542	2495	QWSSSSRHSNISDWSEGISSPPTSMPSQITHIPEAFK	2531			
55	gi 112074	2495	QWSSSSRHSNISDWSEGISSPPTSMPSQITHIPEAFK	2531			

Tables 33E-I list the domain descriptions from DOMAIN analysis results against NOV33. This indicates that the NOV33 sequence has properties similar to those of other proteins known to contain this domain.

Table 33E Domain Analysis of NOV33

gnl|Smart|smart00004, NL, Domain found in Notch and Lin-12; The Notch protein is essential for the proper differentiation of the Drosophila ectoderm. This protein contains 3 NL domains. (SEQ ID NO:825)
CD-Length = 39 residues, 100.0% aligned
Score = 45.1 bits (105), Expect = 5e-05

NOV33: 1443 PPLIEEACELPECQEDAGNKVCSLQCNNHACGWDGGDCS 1481
Sbjct: 1 PQDPWSRCEDAQCWDKFGDGVCDCECNNAECLWDGGDCS 39

Table 33F Domain Analysis of NOV33

gnl|Smart|smart00004, NL, Domain found in Notch and Lin-12; The Notch protein is essential for the proper differentiation of the Drosophila ectoderm. This protein contains 3 NL domains. (SEQ ID NO:825)
CD-Length = 39 residues, 94.9% aligned
Score = 44.7 bits (104), Expect = 7e-05

NOV33: 1486 DPWKNTQSLQCWKYFSDGHCDSSQCNSAGCLFDGFDCC 1523
Sbjct: 3 DPWSRCE-DAQCWDKFGDGVCDCECNNAECLWDGGDCS 39

Table 33G Domain Analysis of NOV33

gnl|Smart|smart00004, NL, Domain found in Notch and Lin-12; The Notch protein is essential for the proper differentiation of the Drosophila ectoderm. This protein contains 3 NL domains. (SEQ ID NO:825)
CD-Length = 39 residues, 97.4% aligned
Score = 41.2 bits (95), Expect = 7e-04

NOV33: 1522 CQRAEGQCNPPLYDQYCKDHFSDGHCDQGCNSAECEWDGLDC 1562
Sbjct: 1 PQDPWSRCE---DAQCWDKFGDGVCDCECNNAECLWDGGDC 38

Table 33H Domain Analysis of NOV33

gnl|Pfam|pfam00023, ank, Ank repeat. Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure. (SEQ ID NO:826)
CD-Length = 33 residues, 97.0% aligned
Score = 42.7 bits (99), Expect = 3e-04

NOV33: 1929 GETALHLAARYSRSDAAKRLLEASADANIQDN 1960
Sbjct: 2 GNTPLHLAARNHLEVVKLLLEAGADVNRDK 33

Table 33I Domain Analysis of NOV33

gnl|Pfam|pfam00066, notch, Notch (DSL) domain. The Notch domain is also called the 'DSL' domain. The notch proteins are transmembrane proteins with extracellular domains of repeated EGF domains and the notch (or DSL) domain N-terminal to that. These proteins are generally involved in lateral inhibition in developmental processes. (SEQ ID NO:826)
 CD-Length = 38 residues, 81.6% aligned
 Score = 42.0 bits (97), Expect = 4e-04

NOV33: 1533 YDQYCKDHFSDGHCDQGCNSAECEWDGLDCA 1563
 | ++| + |++| |+| ||+| | +|| ||+
 Sbjct: 8 YRRHCAERFANGVCNQECNNAACGFDGGDCS 38

5

Notch is a surface receptor. It transmits signals received from outside the cell to the cell's interior. Notch ligands, such as Delta, Serrate and Scabrous interact with epidermal growth factor repeats contained in Notch's extracellular domain. Notch plays an active role in the differentiation of glial cells and Notch influences the length and organisation of neuronal processes. Several homologs of the Drosophila Notch receptor and its ligands, Delta/Serrate, have been cloned in man. Three human disorders including a neoplasia (a T-cell acute lymphoblastic leukemia/lymphoma), a late onset neurological disease (CADASIL) and a developmental disorder (the Alagille syndrome) are associated with mutations in, respectively, the Notch1, Notch3 and Jagged1 genes, pointing out the broad spectrum of Notch activity in humans (Joutel A, and Tournier-Lasserre E, 1998, Semin Cell Dev Biol, 9:619-25; Frisen J, and Lendahl U, 2001, Bioessays 23:3-7).

In Drosophila, the intracellular domain of Notch binds Suppressor of hairless, a multifunction transcription factor that acts as a signal transducing molecule shuttling between the cytoplasm and the nucleus. A nuclear function has been documented for the mammalian Notch homolog (Lu, 1996), as well as for Drosophila Notch (Struhl and Adachi, 1998, Cell 93:649-60). When Notch is bound by a ligand, a signal is passed across the cell membrane releasing the Suppressor of Hairless protein, freeing this protein to enter the nucleus and assume its role in activating transcription of enhancer of Split complex genes. E(spl)-C proteins act in turn to repress the adoption of neural and other differentiated states. Deltex, an intracellular docking protein, replaces Suppressor of Hairless as Su(H) leaves the site of interaction with the intracellular tail of Notch.

The Notch receptor's function is called neurogenic, but this confusing nomenclature refers to the phenotype established in the absence of functional Notch. Notch's function is to repress the adoption of differentiation by cells that carry the Notch protein. A look at the principle ligand of Notch (Delta) and its function makes the anti-neural function of Notch

more easily understood. Delta is not secreted, but is cell bound. The Delta-Notch interaction serves a cell adhesive function between ligand and receptor bearing cells. The receptor bearing cell is inhibited in assuming a differentiated state, while the ligand bearing cell is free to do so. During neurogenesis, this latter cell delaminates, that is, it migrates out of the epithelial cell layer in which it formerly resided, and assumes the differentiated state of a neuroblast in its new physical location within the developing nervous system. Thus Notch is involved in neurogenesis with respect to cells that bears the ligands for Notch: Delta, Serrate and Scabrous.

Lateral inhibition is one of the major themes of development. The process of lateral inhibition and cell selection is repeated hundreds of times in *Drosophila*, with differentiation that takes place in nearly every kind of tissue. For example, Notch is required to limit the number of neuronal precursors, limit the number of muscle precursors, limit the growth of malpighian tubules, and regulate the growth of the ovary. Notch also functions as receptor for both Serrate and Delta in organizing the dorsal-ventral boundary of the wing. One important target of Serrate and Notch in this context is wingless (Diaz-Benjumea and Cohen, 1995, Development 121:4215-25). Two extreme models can be envisioned for lateral inhibition. The first implicates the Notch pathway in the choice of a single precursor via a negative feedback loop. This process could be random in some cases. The second model postulates that the precursor is pre-determined by some mechanism other than Notch signaling, and that Notch signaling then serves only to mediate mutual, uniform repression of other cells and ensure development of a single precursor. Studies concerning the physical spacing of precursors for the microchaetes of the peripheral nervous system suggest the existence of a regulatory loop under transcriptional control between Notch and its ligand Delta. Activation of Notch leads to repression of the achaete-scute genes, which are themselves known to regulate transcription of Delta; this regulation may perhaps be direct (Seugnet et al., 1997, Dev Biol. 192:585-98). Neuroblast segregation was studied in embryos lacking both the maternal and the zygotic forms of either Notch or Delta. A seven-up-LacZ marker was used to follow neuralization of 5-2 and 7-4 neuroblast groups. In the absence of Notch signaling, the cells with an equivalence group do not enter the neural differentiation pathway simultaneously. Neuralization within a group is progressive with two or three cells segregating early and several more later. This suggests that neural potential is not evenly distributed among these cells. A requirement for transcriptional regulation of Notch and/or Delta during neuroblast segregation in embryos was tested by providing Notch and Delta ubiquitously at uniform levels. Neuroblast segregation occurs normally under conditions of uniform Notch expression, suggesting that transcriptional

regulation of Notch is not necessary for many aspects of development of the larval CNS and PNS. In particular, it is dispensable both before and after neuroblast segregation, implying that it is not a necessary component of neuroblast segregation, per se. Under conditions of uniform Delta expression, a single neuroblast segregates from each proneural group in 80% of the cases; in the remaining 20%, more than one neuroblast segregates from a single group of cells. Thus transcriptional regulation of Delta is largely dispensable, with only a small percentage of multiple neurons segregating in each cluster. Genes such as *achaete*, *scute*, *extramacrochaete*, and *wingless* could be responsible for local differences in proneural activity. Notch signaling would enable all cells to mutually repress one another; only a cell with an elevated neural potential could overcome this repression (Seugnet et al., 1997, Development 124:2015-25).

The development and patterning of the wing in *Drosophila* relies on a sequence of cell interactions molecularly driven by a number of ligands and receptors. Genetic analysis indicates that a receptor encoded by the Notch gene and a signal encoded by the *wingless* gene play a number of interdependent roles in this process and display very strong functional interactions. At certain times and places, during wing development, the expression of *wingless* requires Notch activity and that of its ligands Delta and Serrate. This has led to the proposal that all the interactions between Notch and *wingless* can be understood in terms of this regulatory relationship. This proposal has been tested by analyzing interactions between Delta- and Serrate-activated Notch signaling and Wingless signaling during wing development and patterning. Cell death caused by expressing dominant negative Notch molecules during wing development cannot be rescued by coexpressing *Nintra*. This suggests that the dominant negative Notch molecules cannot only disrupt Delta and Serrate signaling but can also disrupt signaling through another pathway. One possibility is the Wingless signaling pathway, since the cell death caused by expressing dominant negative Notch molecules can be rescued by activating Wingless signaling. Furthermore, the outcome of the interactions between Notch and Wingless signaling differs when Wingless signaling is activated by expressing either Wingless itself or an activated form of the Armadillo. For example, the effect of expressing the activated form of Armadillo with a dominant negative Notch on the patterning of sense organ precursors in the wing resembles the effects of expressing Wingless alone. This result suggests that signaling activated by Wingless leads to two effects: a reduction of Notch signaling and an activation of Armadillo (Brennan, 1999, Curr Biol, 9:707-10).

Expression of a dominant negative Notch molecule (Extracellular Notch or ECN) throughout the developing wing mimics the effects of loss of Notch function. However, *Nintra* cannot rescue the cell death caused by overexpressing ECN. Since *Nintra* provides constitutive

signaling for Delta and Serrate during wing development and the effects of ECN are mediated by the sequestration of extracellular molecules that can interact with Notch, this suggests that the ECN molecule is sequestering extracellular molecules other than Delta and Serrate and attenuating signaling through another pathway. One candidate pathway is the Wingless signaling pathway, since the cell death caused by expressing the ECN can be rescued by activating Wingless signaling. Therefore, it is possible that the ECN molecule is sequestering the Wingless protein. The possibility that Wingless can bind the extracellular domain of Notch is supported by the following results, in particular, by two observations: first, that some of the deleterious effects of ECN can be suppressed by Wingless, but not Wingless signaling in the form of a constitutively active Armadillo molecule; and second, that this interaction requires specific EGF-like repeats of Notch, namely repeats 17-19 and 24-26 but not 10-12. Evidence for a physical interaction between Notch and Wingless has also been provided recently by Wesley (1999, *Mol Cell Biol.* 19:5743-58) who finds that the Wingless protein is enriched in a biopanning assay designed to identify proteins that interact with the extracellular domain of the Notch protein and that Wingless can be immunoprecipitated with Notch from embryo extracts and cultured cells. These experiments also show that the association of Wingless with Notch requires the integrity of a region of Notch centered around EGF-like repeats 24-26 (Wesley, 1999, *Mol Cell Biol.* 19:5743-58) which these experiments indicate are essential for the interactions that are described between Wingless and ECN during wing development and patterning (Brennan et al., 1999, *Curr Biol.* 9:707-10). The interaction of Wingless and Notch signaling that has been observed might also be important during normal neural development. Wingless and Delta have opposite effects during neurogenesis; Wingless promotes while Delta suppresses the development of sense organs. Various experiments suggest that during the segregation of neural precursors a reduction of Notch signaling in the precursors themselves is as important as the Delta-mediated activation of Notch signaling in the surrounding cells. It is possible that, like the activation of Notch by Delta, the suppression of Notch signaling is an active process mediated by the interaction of Wingless and Dishevelled with Notch. If this were the case, since both Delta and Wingless have spatially and temporally regulated patterns of gene expression, their interactions with Notch could contribute to the well-documented bias in the appearance of precursors from clusters of cells with neural potential. This competitive interaction could also account for the observed increases in Wingless signaling associated with reductions in Notch signaling during lateral inhibition (Brennan et al., 1999, *Curr Biol.* 9:707-10; Brennan et al, 1999, *Dev Biol.* 216:210-29).

10

20

25

30

hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

5 NOV34

A disclosed NOV34 nucleic acid of 935 nucleotides (also referred to as CG56715-01) encoding a novel Olfactory Receptor-like protein is shown in Table 34A. An open reading frame was identified beginning with an ACA codon, which codes for the amino acid Threonine, at nucleotides 2-4 and ending with a TGA codon at nucleotides 932-934. Putative
10 untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 34A, and the start and stop codons are in bold letters.

Table 34A. NOV34 Nucleotide Sequence (SEQ ID NO:135)

<p><u>CACACAAGAAGGCATCTACTTCATCCTCACGGACATCCCTGGATTGAGGCCTCCACATCTGGATCTCCA</u> TCCCGTCTGCTGTCTCTACACCATCTCCATCATGGGCAATACCACCATCCTCACTGTCATTGCGACAGAG CCATCTGTCCACCAGCGCATGTATCTGTTTCTCTCCATGCTGGCCCTGACGGACCTGGGTCTCACCCCTCAC CACCTACCCACAGTCATGCAGCTTCTCTGGTTCAACGTTTCGTAGAATCAGCTCTGAGGCCCGTTTTGCTC AGTTTTTCTTCCTTCATGGATTCTCCTTTATGGAGTCTTCTGTCTCCTGGCTATGTCGGTGAAGTCTAT GTGGCCATCTGCTGTCCCTCCATTATGCCTCCATCCTCACCAATGAAGTCATTGGTAGAACTGGGTAGC CATCATTTGCTGCTGTGTTCTGGCGGTTCTTCCCTCCCTTTTCTTACTCAAGCGACTGCCTTCTGCCACT CCCACCTTCTCTCGCTCCTATTGCCTCCACCAGGATATGATCCGCTGGTCTGTGCTGACATCAGGCTC AACAGCTGGTATGGATTGCTCTTGCTTGTTCATTATTATCGTGGATCCTCTGCTCATTGTGATCTCCTA TACACTTATTCTGAAAAATATCTTGGGCACAGCCACCTGGGCTGAGCGACTCCGTGCCCTCAATAACTGCC TGTCACCATCTGGCTGTCTGGTCTCTACATTCCCATGGTTGGTGTATCTATGACTCATCGCTTTGCC AAGCATGCCTCTCCACTGGTCCATGTTATCATGGCCAATATCTACCTGCTGGCACCCCGGTGATGAACCC CATCATTTACAGTGTAAGAACAAGCAGATCCAATGGGGAATGTTAAATTTCTTCCCTCAAAAATATGC ATTCAAGATTGAG</p>

The NOV34 nucleic acid has 578 of 903 bases (64%) identical to a gb:GENBANK-
15 ID:AF137396|acc:AF137396.2 mRNA from *Homo sapiens* (ubiquilin 3, HOR 5'Beta14,
HOR5'Beta13, HOR5'Beta12, and HOR5'Beta11 genes, complete cds; HOR 5'Beta10 and
HOR5'Beta9 pseudogenes, complete sequence; HOR5'Beta8 and HOR5'Beta7 genes, complete
cds; CHR11ORF1 and amphiphysin pseudogenes, complete sequence; HOR5'Beta6 and
HOR5'Beta5 genes, complete cds; HOR5'Beta4 pseudogene, complete sequence; HOR 5'Beta3
20 genes, complete cds; HOR5'Beta2 pseudogene, complete sequence; and HOR 5'Beta1 gene,
complete cds) ($E = 6.1e^{-50}$).

A disclosed NOV34 polypeptide (SEQ ID NO:136) encoded by SEQ ID NO:135 is
310 amino acid residues and is presented using the one-letter code in Table 34B. Signal P,
Psort and/or Hydropathy results predict that NOV34 contains a signal peptide and is likely to
25 be localized to the plasma membrane with a certainty of 0.6000. The most likely cleavage site
for a NOV34 peptide is between amino acids 36 and 37: IMG-NT.

Table 34B. Encoded NOV34 protein sequence (SEQ ID NO:136)

TQEGIIYFILTDIPGFEASHIWISIPVCCLYTISIMGNTTILTVIRTEPSVHQRMVFLSMLALTDLGLTLTT
LPTVMQLLWFNVRRISSEARFAQFFFLHGFSESSVLLAMSVDCYVAICCPHYASILTNEVIGRTGLAII
CCCVLAVLPSLFLKRLPFCHSHLLSRSYCLHQDMIRLVCADIRLNSWYGFALALFIIIVDPLLIVISYTLI
LKNILGTATWAERLRALNNCLSHILAVLVLYIPMVGVSMTHRFKASPLVHVIMANIYLLAPPVMNPPIYS
VKNKQIQWGLNFLSLKNMHSR

The disclosed NOV34 amino acid sequence has 160 of 296 amino acid residues (54%) identical to, and 210 of 296 amino acid residues (70%) similar to, the 319 amino acid residue ptnr:TREMBLNEW-ACC:AAG41684 protein from *Mus musculus* (Mouse) (MOR 3'BETA4) (E = 6.3e⁻⁸³).

NOV34 is predicted to be expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV34 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 34C.

Table 34C. BLAST results for NOV34

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17472775 ref XP_061808.1 (XM_061808)	similar to MOR 3Beta4 (H. sapiens) [<i>Homo sapiens</i>]	317	298/300 (99%)	298/300 (99%)	e-136
gi 17456767 ref XP_061618.1 (XM_061618)	similar to prostate specific G-protein coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	879	168/289 (58%)	215/289 (74%)	6e-81
gi 17456753 ref XP_061614.1 (XM_061614)	similar to MOR 3Beta4 (H. sapiens) [<i>Homo sapiens</i>]	315	154/294 (52%)	214/294 (72%)	5e-75

gi 17472781 ref XP_061811.1 (XM_061811)	similar to OLFACTORY RECEPTOR 51I2 (HORSBETA12) (H. sapiens) [<i>Homo sapiens</i>]	312	150/290 (51%)	204/290 (69%)	6e-74
gi 11908220 gb AAG41684.1 (AF133300)	MOR 3'Beta4 [<i>Mus musculus</i>]	319	159/294 (54%)	209/294 (71%)	2e-72

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 34D.

Table 34D Clustal W Sequence Alignment

5

- 1) NOV34 (SEQ ID NO:136)
- 2) gi|17472775|ref|XP_061808.1| (XM_061808) similar to MOR 3Beta4 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:472)
- 3) gi|17456767|ref|XP_061618.1| (XM_061618) similar to prostate specific G-protein coupled receptor (H. sapiens) [*Homo sapiens*] (SEQ ID NO:473)
- 4) gi|17456753|ref|XP_061614.1| (XM_061614) similar to MOR 3Beta4 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:474)
- 5) gi|17472781|ref|XP_061811.1| (XM_061811) similar to OLFACTORY RECEPTOR 51I2 (HORSBETA12) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:475)
- 6) gi|11908220|gb|AAG41684.1| (AF133300) MOR 3'Beta4 [*Mus musculus*] (SEQ ID NO:476)

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

gi	17472775	153	CCVLAVLPSLFLKRLPFCHSHLLSRSYCLHQDMIRLVCADIRLNSWYGFALALLIIVD	212
gi	17456767	721	RSVALTFPLPFLKRFPPYCCSPVLSHSYCLHQEVMKLACADMKANSIYCMFVIIVSTVGID	780
gi	17456753	153	RSLGVVLPPTPLLRHYHYCHGNALSHAFCLHQDVLRSLCTDARTNSIYGLCCVVIATLGVD	212
gi	17472781	151	RSFTILFPLPFLIKRLPICRSNVLSHSYCLHPDMMRLACADISINSIYGLFVLVSTFGMD	210
gi	11908220	153	RSSITLITPLIARLAFFPFCSHVLSHSYCLHQDMIRLACADIRFNVIIYGLVLTLLWGMD	212
5				
10				
NOV34		206	PLLIIVISYTLILKNILGTATWAERLRALNNCLSHILAVLVLYIPMVGVSMTHRFAKHASP	265
gi	17472775	213	PLLIIVISYTLILKNILGTATWAERLRALNNCLSHILAVLVLYIPMVGVSMTHRFAKHASP	272
gi	17456767	781	SLLIIFSYALILRTVLSIASRAERFKALNTCVSHICAVLLIFYTPMIGLSVHRFGKQAPH	840
gi	17456753	213	SIFILSYVLILNTVLDIASREOLKALNTCVSHICVVLIFFVPMIGVSMVHRFGKHLSP	272
gi	17472781	211	LFFIFESYVLILRSVMATASREERLALNTCVSHILAVLAFVPMIGVSMVHRFGKHPVC	270
gi	11908220	213	SLGIFVSYVLILHSVLKIASREGRLKALNTCASHICAVLLIFYPMIGLSVHRFAKHSSP	272
15				
20				
NOV34		266	LVHVMANIYLLAPPVMPNPIIYSVKNKQIQWGMN-----	300
gi	17472775	273	LVHVMANIYLLAPPVMPNPIIYSVKNKQIQWGMNINFLSLKNMHSR--	317
gi	17456767	841	LVQVVMGFMYLLFPVMPNPIIYSVKTKQIRDRVTHAFCY-----	879
gi	17456753	273	IVHITMADTYLLLPVMPNPIIYSVVRTKQIRLGILHKFVLRRRF----	315
gi	17472781	271	YTHVLMENVYLFVPPVLPNPIIYSAKTKERRAIFRMFHHIKI----	312
gi	11908220	273	LHHTFMANIYLLVPPVLPNPIIYSVKTKQIREGILHLLCSPKISSITM	319
25				

Table 34E list the domain descriptions from DOMAIN analysis results against NOV34. This indicates that the NOV34 sequence has properties similar to those of other proteins known to contain this domain.

Table 34E Domain Analysis of NOV34

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 41.6 bits (96), Expect = 7e-05

30	NOV40:	36	GNTTILTVIRTEPSVHQRMVFLSMLALTDLGLTTLTTLPTVMQLLWFNVRRISSEARFAQ	95
	Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
35	NOV40:	96	FFFLHGFSEFMESVLLAMSVDCAICCPHYASILTNEVIGRTGLAIICCVLAVLPSL	155
	Sbjct:	61	GALFVVNGYASILLTAISIDRYLAIVHPLRYRRIRT---PRRAKVLILLVWVLALLS-	116
40	NOV40:	156	FLLKRLPFCHSHLLSRSYCLHQDMIRLVCADIRLNSWYGFALALFIIIVDPLLIVISYTL	215
	Sbjct:	117	-----LPPLLFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFLVPLLVILVCYTR	171
45	NOV40:	216	IL-----KNILGTATWAERLRALNNCLSHILAVLVLYIPMVGVSMTHRFAKHASPLV	267
	Sbjct:	172	ILRTLKRKRSQRSRLKRRSSSERKAAKMLLVVVVVFLCWLPHYHIVLLDSLCLLSIWRV	231
	NOV40:	268	HVIMANIYLL---APPVMNPIIY	287
	Sbjct:	232	LPTALLITLWLAYVNSCLNPIIY	254
50				

G-Protein Coupled Receptor (GPCRs) have been identified as an extremely large family of protein receptors in a number of species. At the phylogenetic level they can be

classified into four major subfamilies. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors. They are likely to be involved in the recognition and transduction of various signals mediated by G-Proteins, hence their name G-Protein Coupled Receptors. The human GPCR genes are generally intron-less and belong to four gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large family of GPCRs in a number of species. As members of the GPCR family, these receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Like GPCRs, the ORs they can be expressed in a variety of tissues where they are thought to be involved in recognition and transmission of a variety of signals. The human OR genes are typically intron-less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The NOV34 nucleic acid of the invention encoding a Olfactory Receptor-like protein includes the nucleic acid whose sequence is provided in Table 34A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 34A while still encoding a protein that maintains its Olfactory Receptor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36% of the residues may be so changed.

The NOV34 protein of the invention includes the Olfactory Receptor-like protein whose sequence is provided in Table 34B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 34B while still encoding a protein that maintains its Olfactory Receptor-like activities and

physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 49% of the bases may be so changed.

The NOV34 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders, Cell Shape disorders, Feeding disorders, control of feeding, potential obesity due to over-eating, potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm, adenocarcinoma, lymphoma, prostate cancer, uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease, multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation, Dentatorubro-pallidoluysian atrophy (DRPLA), Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and other diseases, disorders and conditions of the like.

NOV34 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV34 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV35

A disclosed NOV35 nucleic acid of 1102 nucleotides (also referred to as CG56718-01) encoding a novel Olfactory Receptor-like protein is shown in Table 35A. An open reading

frame was identified beginning with an ATG initiation codon at nucleotides 92-94 and ending with a TGA codon at nucleotides 1049-1051. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 35A, and the start and stop codons are in bold letters.

Table 35A. NOV35 Nucleotide Sequence (SEQ ID NO:137)

ATAACTCTACTAACTAAAAGAAAATGTATCTAAATATTTTACATAACACTTAATGTATCACTTCTATAGGA
TGGAAGCCTTTTGGAGGACAAATGCTGCAAAACAGGACACCATGGAAATCCTAAGCAACTCAACATCTAAA
TTTCCAACCTTCTTGTGACCGGCATTCTGGCCTAGAGTCTGCCCATGTCTGGATCTCCATTCTTTCTG
TTGTTTTATGCCATTGCCCTCTCTGGGAACAGCGTGATCCTGTTGTCACTATTACCCAGCAGAGTCTCC
ATGAACCCATGTATTATTTCTCTTCAGGCTATCAGCCACTGATCTGGGCTTGACTGTTTCTTCATTGTCA
ACAACATTAGGTATCCTCTGGTTTGAGGCACGTGAAATCAGTCTATATAGCTGCATTGTCCAGATGTTTTT
TCTTCATGGATTCACTTTTATGGAATCTGGAGTGCTGGTGGCTACAGCCTTTGACCGTTATGTGGCCATCT
GTGACCCTCTGAGGCACACTACCATTCTCACTAATCCAGAATCATTCAAATGGGTCTTCTGATGATTACA
CGTGCTATAGTACTAATATTGCCACTACTTTTGCTCCTTAAGCCTCTCTATTCTGTAGAATGAATGCCCT
TTCTCACTCCTATTGTTACCATCCAGATGTGATTCAATTAGCATGTTTCCAGACATTCGGGCAATAGCATCT
GTGGATTAAATTGATCTCATCTGACCCTGGAATAGATACACCATGCATTGTCTGTCTATATCTTAATT
ATCACTCTGTCTCAGAAATTGCTTCCCTGAAGAATGGCACAAGGTCTTCAGCACCTGTGTCTCCCATGT
GGGAGCAGTTGCTTTCTTCTACATCCACATGCTGAGCCTGTCTTGGTGTATCGCTATGGTCCGTCAGCCC
CCAGAGTAGTCCATTCACTGATGGCTAATGTATACCTGCTTTTACCCCTGTGCTCAACCCCATCATCGAC
AGTGTA AAAACAAACAAATCCGCAAGGCTATGCTCAGTCTGCTGTACAAAATGAACAGACATAGTTTT
ATTTGATACAAACCTGGCATGAATGACTTGCACTGTA

The NOV35 nucleic acid, located on chromosome 11, has 636 of 1006 bases (63%) identical to a gb:GENBANK-ID:AF133300|acc:AF133300.2 mRNA from *Mus musculus* (MOR 3'Beta1, MOR 3'Beta2, MOR 3'Beta3, and MOR 3'Beta4 genes, complete cds; Cbx3 pseudogene, complete sequence; and MOR 3'Beta5 and MOR 3'Beta6 genes, complete cds) ($E = 1.2e^{-51}$).

A disclosed NOV35 polypeptide (SEQ ID NO:138) encoded by SEQ ID NO:137 is 319 amino acid residues and is presented using the one-letter code in Table 35B. Signal P, Psort and/or Hydropathy results predict that NOV35 contains a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000.

Table 35B. Encoded NOV35 protein sequence (SEQ ID NO:138)

MLQNQDTMEILSNSTSKFPTFLLTGIPGLESABHVISIPFCCFYAIALSGNSVILFVITQQSLHEPMMYYFL
FRLSATDLGLTVSSLSTTLGILWFEAREISLYSCIVQMFFLHGFTFMESGVLVATAFDYVAICDPLRHTTI
LTNSRI IQMGLLMITRAIVLILPLLLLLKPLYFCRMNALSHSYCYHPDVIQLACSDIRANSICGLIDLILTT
GIDTPCIVLSYILIIHSLVLRASPEEWHKVFSTCVSHVGAVAFYIHMLSLSLVYRYGRSAPRVVHSMANV
YLLLPVLPNIIDSVKTKQIRKAMLSLLLT

The disclosed NOV35 amino acid sequence has 160 of 309 amino acid residues (51%) identical to, and 219 of 309 amino acid residues (70%) similar to, the 321 amino acid residue ptnr:TREMBLNEW-ACC:AAG42364 protein from *Homo sapiens* (Human) (odorant receptor HOR3'BETA1) ($E = 2.5e^{-81}$).

NOV35 is predicted to be expressed in at least Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus

callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV35 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 35C.

Table 35C. BLAST results for NOV35					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17456801 ref XP_061626.1 (XM_061626)	similar to OLFACTORY RECEPTOR 51I2 (HORSBETA12) (H. sapiens) [<i>Homo sapiens</i>]	342	196/312 (62%)	251/312 (79%)	1e-95
gi 17456777 ref XP_061621.1 (XM_061621)	similar to olfactory receptor-like protein COR3beta (H. sapiens) [<i>Homo sapiens</i>]	327	165/312 (52%)	218/312 (68%)	6e-77
gi 17456767 ref XP_061618.1 (XM_061618)	similar to prostate specific G-protein coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	879	158/304 (51%)	218/304 (70%)	8e-74
gi 17472781 ref XP_061811.1 (XM_061811)	similar to OLFACTORY RECEPTOR 51I2 (HORSBETA12) (H. sapiens) [<i>Homo sapiens</i>]	312	155/295 (52%)	209/295 (70%)	1e-73
gi 18202242 sp O88628 OXE2_RAT	Olfactory receptor 51E2 (G- protein coupled receptor RA1c)	320	147/305 (48%)	206/305 (67%)	4e-73

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 35D.

Table 35D Clustal W Sequence Alignment

1) NOV35 (SEQ ID NO:138)		
5	2) gi 17456801 ref XP_061626.1 (XM_061626) similar to OLFACTORY RECEPTOR 51I2 (HOR5BETA12) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:477)	
	3) gi 17456777 ref XP_061621.1 (XM_061621) similar to olfactory receptor-like protein COR3beta (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:478)	
	4) gi 17456767 ref XP_061618.1 (XM_061618) similar to prostate specific G-protein coupled receptor (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:479)	
	10 5) gi 17472781 ref XP_061811.1 (XM_061811) similar to OLFACTORY RECEPTOR 51I2 (HOR5BETA12) (H. sapiens) [<i>Homo sapiens</i>] (SEQ ID NO:480)	
	6) gi 18202242 sp O88628 OXE2_RAT Olfactory receptor 51E2 (G-protein coupled receptor RA1c) (SEQ ID NO:481)	
15		10 20 30 40 50 60
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
20	gi 17456767 1	MSLALDLCPLSQRL EAFSSIVLFFQTAPAVRHPKGLLELHKTVP TSIKEELKGFPTSD 60
	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
25		70 80 90 100 110 120
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456767 61	HFIIITDFIAKYHTDLKWAVLGIATPRQQFKALNTCISHICAVLIFYVPTLSAAMLHQFAR 120
30	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
35		130 140 150 160 170 180
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456767 121	DVSPMIHVLMA DIFLLVPPLLNPIVYCVKTHQIREKVVGKLC PKNCF LKSKILPRCSFVP 180
	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
40	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
45		190 200 210 220 230 240
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456767 181	GFRLAYYYLP PPHPSKVSFLDPVEKANRSAPTQFSPMP SADASLLADLGT FSSLQRATFFL 240
	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
50		250 260 270 280 290 300
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456767 241	TGFOGLEGLHGWISIPFCFIYLT VILGNLTILHVICTDATLHGPMYYFLGMLAVTD LGLC 300
	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
55		310 320 330 340 350 360
	NOV35 1
	gi 17456801 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456777 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
	gi 17456767 301	LSTLPTVLGIFWFD TREIGIPACFTQLFFIHTLSSMESSVLLSMSIDRYVAVCNPLHDST 360
	gi 17472781 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
60	gi 18202242 1	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
65		


```

gi|18202242| 210 VMFISLSYFLITRAVLQIPSKSERAKAFGTCVSHIGVVLAFYVPLICLSVVHRECNLSLDP 269

      850      860      870      880      890      900
5  NOV35      280 VVHSVMANVYLLPPVLPNIIDSVTKQIRK---AMLSLLLT----- 319
   gi|17456801| 285 FVHITMANVFLPPVLPNPIIYSVKIKQIQK---AIKVLIOKHSKSNHQLFLIRDKAIY 341
   gi|17456777| 273 VLCSTLANIYLLPPVLPNPIIYSLKTKIIRQ---AMFQLQSKGSWGFNVRGLRGRWD-- 327
   gi|17456767| 841 LVQVVMGFMVYLLPPVLPNPIIYSVKTKQIRDR---VTHAFCY----- 879
   gi|17472781| 271 YIHVLMENVYLFVPPVLPNPIIYSAKTKEIRR---ALFRMPHHIKI----- 312
10 gi|18202242| 270 IVHVMCDVYLLPPVLPNPIIYGAKTKQIRTRVLAMFKISCDKDIEAGGNT----- 320

NOV35      319 - 319
15 gi|17456801| 342 E 342
   gi|17456777| 327 - 327
   gi|17456767| 879 - 879
   gi|17472781| 312 - 312
   gi|18202242| 320 - 320
20

```

Table 35E list the domain descriptions from DOMAIN analysis results against NOV35. This indicates that the NOV35 sequence has properties similar to those of other proteins known to contain this domain.

Table 35E Domain Analysis of NOV35

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family)
 (SEQ ID NO:810)
 CD-Length = 254 residues, 40.9% aligned
 Score = 56.2 bits (134), Expect = 3e-09

```

NOV35: 50  GNSVILFVIITQOSLHEPMYFYFLRSLATDLGLTVSSLSTTLGILWFEAREISLYSCIVQ 109
          || +++ ||+ + | | || |+ || ++ | | | +
Sbjct: 1  GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV 60

NOV35: 110 MFFLHGFTFMESGVLVATAFDRYVAICDPLRHTTILNSRIIQM 153
          + + | | + |||+|| |||+ | | | +
Sbjct: 61 GALFVVNGYASILLTLAISIDRYLAIVHPLRYRRIRTPRRAKVL 104

```

G-Protein Coupled Receptor (GPCRs) have been identified as an extremely large family of protein receptors in a number of species. At the phylogenetic level they can be classified into four major subfamilies. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors. They are likely to be involved in the recognition and transduction of various signals mediated by G-Proteins, hence their name G-Protein Coupled Receptors. The human GPCR genes are generally intron-less and belong to four gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large family of GPCRs in a number of species. As members of the GPCR family, these receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are

likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Like GPCRs, the ORs they can be expressed in a variety of tissues where they are thought to be involved in recognition and transmission of a variety of signals. The human OR genes are typically intron-less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The NOV35 nucleic acid of the invention encoding a Olfactory Receptor-like protein includes the nucleic acid whose sequence is provided in Table 35A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 35A while still encoding a protein that maintains its Olfactory Receptor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37% of the residues may be so changed.

The NOV35 protein of the invention includes the Olfactory Receptor-like protein whose sequence is provided in Table 35B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 35B while still encoding a protein that maintains its Olfactory Receptor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 52% of the bases may be so changed.

The NOV35 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders, Cell Shape disorders, Feeding disorders, control of feeding, potential obesity due to over-eating, potential disorders due to

starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm, adenocarcinoma, lymphoma, prostate cancer, uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease, multiple sclerosis, treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation, Dentatorubro-pallidoluysian atrophy (DRPLA), Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and other diseases, disorders and conditions of the like.

NOV35 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV35 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV36

NOV36 includes two novel cadherin 11-like proteins disclosed below. The disclosed sequences have been named NOV36a and NOV36b.

NOV36a

A disclosed NOV36a nucleic acid of 2476 nucleotides (also referred to as CG56729-01) encoding a novel cadherin 11-like protein is shown in Table 36A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 46-48 and ending with a TGA codon at nucleotides 2389-2391. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 36A, and the start and stop codons are in bold letters.

Table 36A. NOV36a Nucleotide Sequence (SEQ ID NO:139)

TCAGCACCCAGGGCCAGTGAACAGAGCCCTGGCTGGAGTCCAAACATGTGGGGCCTGGTGAGGCTCCTGC
TGGCCTGGCTGGGTGGCTGGGGCTGCATGGGGCTCTGGCAGCCCCAGCCCGGCTGGGCAGGGTCCCG
GGAACACCCAGGGCCTGCTCTGCTGCGGACTCGAAGGAGCTGGGTCTGGAACAGTTCTTTGTCAATTGAG
GAATATGCTGGTCCAGAGCCTGTTCTCATTTGCAAGCTGCACTCGGATGTTGACCGGGGAGAGGGCCGCA
CCAAGTACCTGTTGACCGGGGAGGGGGCAGGCACCGTATTGTGATTGATGAGGCCACAGGCAATATTCA
TGTTACCAAGAGCCTTGACCGGGAGGAAAGGCGCAATATGTGCTACTGGCCCCAAGCCGTGGACCGGAGCC
TCCAACCGGGCCCTGGAGCCCCCATCAGAGTTCATCATCAAAGTGAAGACATCAACGACAATCCACCCA
TTTTTCCCCTTGGGGCCTACCATGCCACCGTGCCCGAGATGTCCAATGTCGGTACATCAGTGATCCAGGT
GACTGCTCACGATGCTGATGACCCAGCTATGGGAACAGTGCCAAGCTGGTGTACACTGTTCTGGATGGA
CTGCTTTTCTTCTCTGTGGACCCCCAGACTGGTGTGGTGGTACAGCCATCCCCAACATGGACCGGAGAG
CACAGGAGGAGTTCTTGGTGGTATCCAGGCCAAGGACATGGGCGGCCACATGGGGGGCTGTCAAGCAG
CACTACGGTGACTGTCACGCTCAGCGATGTCAACGACAACCCCCCAAGTTCCACAGAGTCTATACCAG
TTCTCCGTGGTGGAGACAGCTGGACCTGGCACACTGGTGGGCGGGCTCCGGGCCAGGACCCAGACCTGG
GGGACAACGCCCTGATGGCATAACAGCATCCTGGATGGGGAGGGGTCTGAGGCCCTTACAGTACAGACAGA
CTTGACAGGTCGAGACGGGCTCCTCACTGTCCGCAAGGTTCTAGACTTTGAGAGCCAGCGCTCCTACTCC
TTCGTGTCGAGGCCACCAACACGCTCATTTGACCCAGCCATCTGCGGCGAGGGCCCTTCAAGGATGTGG
CCTCTGTGCGTGTGGCAGTGCAAGATGCCCCAGAGCCACCTGCCTTACCCAGGTGCCTACCACCTGAC
AGTGCTGAGAACAAAGGCCCGGGGACCCCTGGTAGGCCAGATCTCCGCGGCTGACTGGACTCCCCCTGCC
AGCCCAATCAGGTACTCCATCCTCCCCACTCAGATCCGGAGCGTTGCTTCTCTATCCAGCCCGAGGAAG
GCACCATCCATACAGCAGCACCCTGGATCGCGAGGCTCGCGCCTGGCACAACCTCACTGTGCTGGCTAC
AGAGCTCGGTGAGGACTCCAGGCCCTCGCGCTGCAAGTGCCCATCCAGACCCTGGATGAGAATGACAAT
GCTCCCCAGCTGGCTGAGCCCTACGATACTTTTGTGTGTGACTCTGCAGCTCCTGGCCAGCTGATTACAG
TCATCCGGGCCCTGGACAGAGATGAAGTTGGCAACAGTAGCCATGTCTCCTTTCAAGGTCCTCTGGGGCC
TGATGCCAACTTTACTGTCCAGGACAACCGAGATGGCTCCGCCAGCCTGCTGCTGCCCTCCCGCCCTGCT
CCACCCCGCATGCCCCCTACTTGGTTCCCATAGAAGCTGTGGGACTGGGGGCAGCCGGCGCTGAGCAGCA
CTGCCACAGTGACTGTTAGTGTGTGCCGCTGCCAGCCTGACGGCTCTGTGGCATCTGCTGGCCTGAGGC
TCACCTCTCAGCTGCTGGGCTCAGCACCGGCGCCCTGCTTCCCATCATCAGCTGTGGGTGCCCTGCTT
GCCCTGGTGGTGTCTTCTGTTGGCCCTGCGGCGGCGAGAAGCAAGAAGCACTGATGGTACTGGAGGAGGAG
ACGTCGAGAGAACATCATCACCTACGACACGAGGGCGGCGGCGAGGAGGACACCGAGGCCCTTCGACAT
CACGGCCTTGACAGAACCCGGACGGGGCGGCCCCCGGCGCCCGGCCCTCCCGCGGCGGAGACGTGTTG
CCCCGGGCGCGGTGTGCGCCAGCCAGACCCCGGCCCGCCGACCTGGGCGCAGCTCCTGGCGCTGC
GGCTCCGCGAGGCGGACGAGGACCCCGCGGTACCCCGTACGACTCGGTGAGGTGTAGDPLDGNALMAYSID
CGGCGGCTCCTCTTGGGCTCCTCAGCTCCTGGGCTCCGGCAGCGAAGCCGGCGGCGCCCCCGGCC
GCGGAGCCGCTGGACGACTGGGGTCCGCTCTTCCGACCCCTGGCCGAGCTGTATGGGGCCAAGGAGCCCC
CGGCCCCCTGAGCGCCCGGCTGGCCCGGCCACCGCGGGGGGGGGGAGCGGGACAGGCCCTCTGAGT
GAGCCCCACGGGTCCAGGCGGGCGG

The disclosed NOV36a nucleic acid sequence, located on chromosome 14, has 992 of 1514 bases (65%) identical to a gb:GENBANK-ID:HUMCA11A|acc:L34056.1 mRNA from *Homo sapiens* (cadherin-11 mRNA, complete cds) ($E = 7.3e^{-145}$).

- 5 A disclosed NOV36a polypeptide (SEQ ID NO:140) encoded by SEQ ID NO:139 is 781 amino acid residues and is presented using the one-letter amino acid code in Table 36B. Signal P, Psort and/or Hydropathy results predict that NOV36a contains a signal peptide and is likely to be localized in the mitochondrial inner membrane with a certainty of 0.8227 in one embodiment and to the plasma membrane with a certainty of 0.4400 in an additional
- 10 embodiment. The most likely cleavage site for a NOV36a polypeptide is between amino acids 16 and 17: GWG-CM.

Table 36B. Encoded NOV36a protein sequence (SEQ ID NO:140).

MWGLVRLLLAWLGGWCMGRLAAPARAWAGSREHPGALLRTRRSWVWNQFFVIEEYAGPEPVLIQKLHSDVDRG
EGRTKYLTTGEGAGTVFVIDEATGNIHVTKSLDREEKAQYVLLAQAVDRASNRPLEPPSEFIKVQDINDNPPIF
PLGPYHATVPMSNVGTSVIQVTAHDADDPYSGNSAKLVYTVLDGLPFFSVDPQTGVVRTAIPNMDRETQEEFLV
VIQAKDMGGHMGGLSGSTTVTVTLSDVNDNPPKFPQSLYQFSVVETAGPGLVGRRLAQDPLDGNALMAYSID
GEGSEAFSISTDLQGRDGLLTVRKVLDFESQRSYSFRVEATNTLIDPAYLRRGPFDVASVRVAVQDAPEPPAFT
QAAYHLTVPENKAPGTLVGQISAADLDPASPIRYSILPHSDPERCFISIQPEEGTHTAAPLDREARAWNLTVL
ATELGEDSQASRVQVAIQTLDENDNAPQLAEPYDTFVCDASAPGQLIQVIRALDRDEVGNSSHVSFQGPLGPDAN

FTVQDNRDGSASLLLPSPAPPRHAPYLVPIELWDWGQPALSSSTATVTVSVCRCPDGSVASWCWPEAHLAAGLS
 TGALLAIITCVGALLALVVLFFVALRRQKQEALMVLEEDVRENIITYDDEGGGEEDTEAFDITALQNPDAAPPA
 PGPPARRDVLPRARVSRQPRPPGPADVAQLLALRLREADEDPGVPPYDSVQVYGYEGRGSSCGSLSSLSGSGSEAG
 GAPGPAEPLDDWGFLFRTLAEYGAKEPPAP

The disclosed NOV36a amino acid sequence has 434 of 746 amino acid residues (58%) identical to, and 552 of 746 amino acid residues (73%) similar to, the 796 amino acid residue ptnr:SWISSPROT-ACC:P55287 protein from *Homo sapiens* (Human) (CADHERIN-11 precursor (osteoblast-cadherin) (OB-cadherin) (OSF-4)) ($E = 2.3e^{-229}$).

NOV36a is predicted to be expressed in at least Cerebral Medulla/Cerebral white matter, Gall Bladder, Retina, Temporal Lobe and Uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV36a is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HUMCA11A|acc:L34056.1) a closely related *Homo sapiens* cadherin-11 mRNA, complete cds homolog in species *Homo sapiens*: osteoblasts.

NOV36b

A disclosed NOV36b nucleic acid of 2476 nucleotides (also referred to as CG56729-02) encoding a novel Cadherin 11-like protein is shown in Table 36C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 46-48 and ending with a TGA codon at nucleotides 2389-2391. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 36C, and the start and stop codons are in bold letters.

Table 36C. NOV36b Nucleotide Sequence (SEQ ID:NO:141)

TCAGCACCCAGGGCCAGTGAACAGAGCCCTGGCTGGAGTCCAAACAT**GTGGGGCCTGGTGAGGCTCCTGC**
TGGCCTGGCTGGGTGGCTGGGGCTGCATGGGGCGTCTGGCAGCCCCAGCCCGGGCCTGGGCAGGGTCCCG
GGAACCCAGGGCCTGCTCTGCTGCGGACTCGAAGGAGCTGGGTCTGGAACAGTTCTTTGTCAATTGAG
GAATATGCTGGTCCAGAGCCTGTTCTCATTGGCAAGCTGCACTCGGATGTTGACCGGGGAGAGGGCCGCA
CCAAGTACCTGTTGACCGGGGAGGGGGCAGGCACCGTATTTGTGATTGATGAGGCCACAGGCAATATTCA
TGTTACCAAGAGCCTTGACCGGGAGGAAAAGGCGCAATATGTGCTACTGGCCCAAGCGTGGACCGAGCC
TCCAACCGGCCCCCTGGAGCCCCCATCAGAGTTCATCATCAAAGTGCAAGACATCAACGACAATCCACCCA
TTTTTCGGCTTGGGCCCTACCATGCCACGTGCCCGAGATGTCCAATGTCGGTACATCAGTGATCCAGGT
GACTGCTCACGATGCTGATGACCCAGCTATGGGAACAGTGCCAAGCTGGTGTACACTGTTCTGGATGGA
CTGCCTTTCTTCTGTGGACCCCCAGACTGGTGTGGTGCCTACAGCCATCCCCAACATGGACCGGGAGA
CACAGGAGGAGTTCTTGGTGGTGATCCAGGCCAAGGACATGGGCGGCCACATGGGGGGGCTGTGAGGCAG
CACTACGGTGACTGTACGCTCAGCGATGTCAACGACAACCCCCCAAGTTCCACAGAGTCTATACAG
TTCTCCGTGGTGGAGACAGCTGGACCTGGCACACTGGTGGGCGGCTCCGGGCCCAGGACCCAGACCTGG
GGGACAACGCCCCTGATGGCATAACAGCATCCTGGATGGGGAGGGGTCTGAGGCCCTTCAGCATCAGCACAGA
CTTGCAGGGTCGAGACGGGCTCCTCACTGTCCGTAAGCCCCCTAGACTTTGAGAGCCAGCGCTCCTACTCC
TTCCGTGTCGAGGCCACCAACAGCTCATTGACCCAGCCTATCTGCGGCGAGGGCCCTCAAGGATGTGG
CCTCTGTGCGTGTGGCAGTGCAAGATGCCCCAGAGCCACCTGCCTTCACCCAGGCTGCCTACCACCTGAC
AGTGCCCTGAGAACAAGGCCCGGGGACCTGGTAGCCAGATCTCCGCGGCTGACCTGGACTCCCCCTGCC

AGCCCAATCAGGTACTCCATCCTCCCCACTCAGATCCGAGCGTTGCTTCTCTATCCAGCCCAGGAAAG
GCACCATCCATACAGCAGCACCCCTGGATCGCGAGGCTCGCGCTGGCACAACCTCACTGTGCTGGCTAC
AGAGCTCGGTGAGGACTCCCAGGCCCTCGCGCTGCAAGTGGCCATCCAGACCCCTGGATGAGAATGACAAT
GCTCCCCAGCTGGCTGAGCCCTACGATACCTTTGTGTGTGACTCTGCAGCTCCTGGCCAGCTGATTACAGG
TCATCCGGGCCCTGGACAGAGATGAAGTTGGCAACAGTAGCCATGTCTCCTTCAAGGTCTCTGGGCCC
TGATGCCAACTTTACTGTGCCAGACAACCGAGATGGCTCCGCCAGCCTGCTGTGCCCTCCCGCCCTGCT
CCACCCCGCCATGCCCCCTACTTGGTTCCCATAGAAGTGTGGGACTGGGGGAGCCGGCGCTGAGCAGCA
CTGCCACAGTGACTGTTAGTGTGTGCCGCTGCCAGCCTGACGGCTCTGTGGCATCCTGCTGGCCTGAGGC
TCACCTCTCAGCTGCTGGGCTCAGCACCGGCGCCCTGCTTGCCATCATCACCTGTGTGGGTGCCCTGCTT
GCCCTGGTGTGCTCTTCGTGGCCCTGCGGCGGCAAGAAGCAAGAAGCACTGATGGTACTGGAGGAGGAGG
ACGTCCGAGAGAACATCATCACCTACGACGACGAGGGCGGCGGAGGAGGACACCGAGGCCTTCGACAT
CACGGCCTTGCGAGAACCCGGACGCGGCGGCCCGGCCCGGCCCTCCCGCGCGCCGAGACGTGTTG
CCCCGGGCCCCGGGTGTGCGGCCAGCCAGACCCCCCGGCCCGGCCGACGTGGCGCAGCTCTGGCGCTGC
GGCTCCGCGAGGCGGACGAGGACCCCGGCGTACCCCGGTACGACTCGGTGCAGGTGTACGGTACGAGGG
CCGCGGCTCCTCTTGCGGCTCCCTCAGCTCCCTGGGCTCCGGCAGCGAAGCCGGCGGCGCCCCCGGCC
GCGGAGCCGCTGGACGACTGGGGTCCGCTCTTCGCAACCTGGCCGAGCTGTATGGGGCCAAGGAGCCCC
CGGCCCTTGAGCGCCCGGCTGGCCCGGCCACCGCGGGGGGGGGGAGCGGGACAGGCCCTCTGAGT
GAGCCCCACGGGGTCCAGCGGGCGG

The disclosed NOV36b nucleic acid sequence, located on chromosome 11, has 1100 of 1109 bases (99%) identical to a gb:GENBANK-ID:AK025342|acc:AK025342.1 mRNA from *Homo sapiens* (cDNA: FLJ21689 fis, clone COL09459) ($E = 2.3e^{-240}$).

A disclosed NOV36b polypeptide (SEQ ID NO:142) encoded by SEQ ID NO:141 is 781 amino acid residues and is presented using the one-letter amino acid code in Table 36D. Signal P, Psort and/or Hydropathy results predict that NOV36b contains a signal peptide and is likely to be localized to the mitochondrial inner membrane with a certainty of 0.8227 in one embodiment and to the plasma membrane with a certainty of 0.4400 in an additional embodiment. The most likely cleavage site for a NOV36b peptide is between amino acids 16 and 17: GWG-CM.

Table 36D. Encoded NOV36b protein sequence (SEQ ID NO:142).

MWGLVRLLLAWLGGWGCMLRLAAPARAWAGSREHPGALLRTRRSWVWNQFFVIEEYAGPEPVLIGKLHSDVDRG
EGRTKYLLTGEGAGTVFVIDEATGNIHVTKSLDREEKAQVLLAQAVDRASNRPLEPPSEFIKQVDINDNPPIF
PLGPYHATVPEMSNVGTSVIQVTAHDADDPYSGNSAKLVYTVLDGLPFFSVDPQTGVVVRTAIPNMDRETQEEFLV
VIQAKDMGGHMGGLSGSTTVTVTLSDVNDNPPKFPQSLYQFSVVETAGPGLVGRRLAQDPDLGDNALMAYSILD
GEGSEAFSISTDLQGRDGLLTVRKPLDFESQRSYSFRVEATNTLIDPAYLRRGPFKDVASVRVAVQDAPEPPAFT
QAAYHLTVPENKAPGTLVGQISAADLDSPASPIRYSILPHSDPERCFSIQPEEGTIHTAAPLDREARAWNLTVL
ATELGEDSQASRVQVAIQTLDENDNAPQLAEPYDTFVCDSPAAPGQLIQVIRALDRDEVGNSSHVFSQGLGPDAN
FTVQDNRDGSASLLPSRPAPPRHAPYLVPIELWDWGQPALSSSTATVTVSVCRCPDGSVASCWPEAHLASAAGLS
TGALLAIITCVGALLALVVLFFVALRRQKQEALMVLEEDVRENIITYDDEGGGEEDTEAFDITALQNPDAAPPA
PGPPARRDVLPRARVSRQPRPPGADVAQLLRLREADEDGVPVPPYDSVQVYGYEGRGSSCGSLSSLSGSGSEAG
GAPGPAEPLDDWGFLFRTLAEELYGAKEPPAP

The disclosed NOV36b amino acid sequence has 435 of 746 amino acid residues (58%) identical to, and 553 of 746 amino acid residues (74%) similar to, the 796 amino acid residue ptnr:SWISSNEW-ACC:P55287 protein from *Homo sapiens* (Human) (cadherin-11 precursor (osteoblast-cadherin) (OB-cadherin) (OSF-4)) ($E = 2.5e^{-230}$).

NOV36b is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia

In addition, NOV36b is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK025342|acc:AK025342.1) a closely related *Homo sapiens* cDNA: FLJ21689 fis, clone COL09459 homolog in species *Homo sapiens*: osteoblasts.

NOV36a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 36E.

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16553903 dbj BAB71613.1 (AK057922)	unnamed protein product [<i>Homo sapiens</i>]	493	459/466 (98%)	461/466 (98%)	0.0
gi 13626134 sp O93319 CADB_CHICK	CADHERIN-11 PRECURSOR	792	430/749 (57%)	552/749 (73%)	0.0
gi 3377485 gb AAC28073.1 (AF002983)	cadherin precursor [<i>Xenopus laevis</i>]	794	429/751 (57%)	547/751 (72%)	0.0
gi 1705549 sp P55288 CADB_MOUSE	CADHERIN-11 PRECURSOR (OSTEOBLAST- CADHERIN) (OB- CADHERIN) (OSF-4)	796	432/753 (57%)	552/753 (72%)	0.0
gi 1377894 dbj BAA04798.1 (D21254)	OB-cadherin-1 [<i>Homo sapiens</i>]	796	434/753 (57%)	552/753 (72%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 36F.

Table 36F Information for the ClustalW proteins

- 1) NOV36a (SEQ ID NO:140)
- 2) NOV36b (SEQ ID NO:142)
- 3) gi|16553903|dbj|BAB71613.1| (AK057922) unnamed protein product [*Homo sapiens*] (SEQ ID NO:482)
- 4) gi|13626134|sp|O93319|CADB_CHICK CADHERIN-11 PRECURSOR (SEQ ID NO:483)
- 5) gi|3377485|gb|AAC28073.1| (AF002983) cadherin precursor [*Xenopus laevis*] (SEQ ID NO:484)
- 6) gi|1705549|sp|P55288|CADB_MOUSE CADHERIN-11 PRECURSOR (OSTEOBLAST-CADHERIN) (OB-CADHERIN) (OSF-4) (SEQ ID NO:485)
- 7) qi|1377894|dbj|BAA04798.1| (D21254) OB-cadherin-1 [*Homo sapiens*] (SEQ ID NO:486)

	Score	Position
NOV36a	1	10 20 30 40 50 60
NOV36b	1	
gi 16553903	1	

Sequence alignment (conserved residues highlighted):
NOV36a: -----MWGLVRILLLAWLGGGWGCGRLLAAPARAWACSRE--HPGPALLRTRRSVVWNQF 51
NOV36b: -----MWGLVRILLLAWLGGGWGCGRLLAAPARAWACSRE--HPGPALLRTRRSVVWNQF 51
gi|16553903: -----MWGLVRILLLAWLGGGWGCGRLLAAPARAWACSRE--HPGPALLRTRRSVVWNQF 51

5

gi 13626134	1	MKEDNCLHAAITCLGMLYSHAITTEKLNHVRPSLHGHHKKGKEGQVLHRSKRGWVWNQF	60
gi 3377485	1	MKKDFCLHGLLCLGLIYCSHATSLRKNKLRQSFHGHHEKKGEGQVLHRSKRGWVWNQF	60
gi 1705549	1	MKENYCLQAALVCLSMLYHSQAFALERRSHLHPSEFHGHHEKKGEGQVLQRSKRGWVWNQF	60
gi 1377894	1	MKENYCLQAALVCLGMLCHSHAFAPERRGHLRPSFHGHHEKKGEGQVLQRSKRGWVWNQF	60

10

NOV36a	52	FVIEEYAGPEPVLIGKLHSDVDRCGEGRTKYLLTGEGAGTVFVIDEATGNIHVTKSLDREE	111
NOV36b	52	FVIEEYAGPEPVLIGKLHSDVDRCGEGRTKYLLTGEGAGTVFVIDEATGNIHVTKSLDREE	111
gi 16553903	52	FVIEEYAGPEPVLIGKLHSDVDRCGEGRTKYLLTGEGAGTVFVIDEATGNIHVTKSLDREE	111
gi 13626134	61	FVIEEYTGPDVPLVGRHLSDIDSQGNIKYILSGEGAGTIFVIDDKSGNIHATKTLDREE	120
gi 3377485	61	FVIEEYTGPDVPLVGRHLSDIDSQGNIKYILSGEGAGTIFVIDDKSGNIHATKTLDREE	120
gi 1705549	61	FVIEEYTGPDVPLVGRHLSDIDSQGNIKYILSGEGAGTIFVIDDKSGNIHATKTLDREE	120
gi 1377894	61	FVIEEYTGPDVPLVGRHLSDIDSQGNIKYILSGEGAGTIFVIDDKSGNIHATKTLDREE	120

15

NOV36a	112	KAQYVLLAQAVDRASNRPLEPPSEFIIKVQDINDNPPFIPLGPHYATVPMSNVGTSVIO	171
NOV36b	112	KAQYVLLAQAVDRASNRPLEPPSEFIIKVQDINDNPPFIPLGPHYATVPMSNVGTSVIO	171
gi 16553903	112	KAQYVLLAQAVDRASNRPLEPPSEFIIKVQDINDNPPFIPLGPHYATVPMSNVGTSVIO	171
gi 13626134	121	RAQYTLTAQAVDRNTNRPLEPPSEFIVKVQDINDNPPFELHENYHANVPERSNVGTSVIO	180
gi 3377485	121	RAQYTLTAQAVDRNTNRPLEPPSEFIVKVQDINDNPPFELHENYHANVPERSNVGTSVIO	180
gi 1705549	121	RAQYTLTAQAVDRNTNRPLEPPSEFIVKVQDINDNPPFELHEIYHANVPERSNVGTSVIO	180
gi 1377894	121	RAQYTLTAQAVDRNTNRPLEPPSEFIVKVQDINDNPPFELHETIYHANVPERSNVGTSVIO	180

20

NOV36a	172	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
NOV36b	172	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
gi 16553903	172	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
gi 13626134	181	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
gi 3377485	181	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
gi 1705549	181	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240
gi 1377894	181	VTASDADDPHYGNSAKLVYSILEGOPYFSVEAQTGII RTALPNMDREAKEEYHVVIQAKD	240

25

NOV36a	232	MGGHMGGLSGSTTTVTITLSDVNDNPPKFPQSLYQFSVVETAGPGTILVGRRLRAQDPDLGDN	291
NOV36b	232	MGGHMGGLSGSTTTVTITLSDVNDNPPKFPQSLYQFSVVETAGPGTILVGRRLRAQDPDLGDN	291
gi 16553903	232	MGGHMGGLSGSTTTVTITLSDVNDNPPKFPQSLYQFSVVETAGPGTILVGRRLRAQDPDLGDN	291
gi 13626134	241	MGGHMGGLSGTTKVTITLSDVNDNPPKFPQSAVPMSVSEAAVPGEEVGRVRAKDPDIDIGEN	300
gi 3377485	241	MGGHMGGLSGTTKVTITLSDVNDNPPKFPQSAVPMSVSEAAVPGEEVGRVRAKDPDIDIGEN	300
gi 1705549	241	MGGHMGGLSGTTKVTITLSDVNDNPPKFPQSAVPMSVSEAAVPGEEVGRVRAKDPDIDIGEN	300
gi 1377894	241	MGGHMGGLSGTTKVTITLSDVNDNPPKFPQSAVPMSVSEAAVPGEEVGRVRAKDPDIDIGEN	300

30

NOV36a	292	ALMAYSILDGEGSEAFSISTDLQGRDGLITVRKVLDFESQRSYSFRVEATNTLIDPAYLR	351
NOV36b	292	ALMAYSILDGEGSEAFSISTDLQGRDGLITVRKVLDFESQRSYSFRVEATNTLIDPAYLR	351
gi 16553903	292	ALMAYSILDGEGSEAFSISTDLQGRDGLITVRKVLDFESQRSYSFRVEATNTLIDPAYLR	351
gi 13626134	301	GLVAYSITDCDGMDFEITTDYETQEGVVKLKKVLDFTKKSYSLKVEAANVHIDPKFIS	360
gi 3377485	301	GLIKYRILEGDCGDMFEITADYVVTQEGVVKLKKVLDYETKKFYSMKVEAANVHIDPKFIS	360
gi 1705549	301	GLVTYNTVDCGIELFEITTDYETQEGVVKLKKVLDFTKKRAYSLKVEAANVHIDPKFIS	360
gi 1377894	301	GLVTYNTVDCGMESEFEITTDYETQEGVVKLKKVLDFTKKRAYSLKVEAANVHIDPKFIS	360

35

NOV36a	352	RGPFKDVASVRVAVQDAPEPPAFTQAAYHLTVPENKAPGTLVGOISAADLDSPASPIRYS	411
NOV36b	352	RGPFKDVASVRVAVQDAPEPPAFTQAAYHLTVPENKAPGTLVGOISAADLDSPASPIRYS	411
gi 16553903	352	RGPFKDVASVRVAVQDAPEPPAFTQAAYHLTVPENKAPGTLVGOISAADLDSPASPIRYS	411
gi 13626134	361	NGPFKDTVTIVKTI VEDAEPPEFLKPSYIHEVOENAAAGTVVGRVHAKDPDAANSPIRYS	420
gi 3377485	361	RGPFKDTATVKISVEDFDEPPIFLERSYILEVYENAPSDTVVGRVHAKDPDAANSPIRYS	420
gi 1705549	361	NGPFKDTVTIVKISVEDADEPPMFLAPSYIHEVOENAAAGTVVGRVHAKDPDAANSPIRYS	420
gi 1377894	361	NGPFKDTVTIVKTI VEDAEPPEMFLAPSYIHEVOENAAAGTVVGRVHAKDPDAANSPIRYS	420

40

NOV36a	412	ILPHSDPERCFSTIOPEEGTHTTAAPLDREARAWHNLTVLATELGEDSASRVQVAIQTLTD	471
NOV36b	412	ILPHSDPERCFSTIOPEEGTHTTAAPLDREARAWHNLTVLATELGEDSASRVQVAIQTLTD	471
gi 16553903	412	ILPHSDPERCFSTIOPEEGTHTTAAPLDREARAWHNLTVLATELGEDSRAHKAAAAS----	467

45

50

55

60

65

70

5

gi | 13626134 | 421 IDRHITDLERYFITINADDGNITIKALDREETIAWHNISVFAVEVHKQHQAERVPVAIKVMD 480
 gi | 3377485 | 421 IDRHITDLDRFFFSINPEDGVIKTKCLDREESFWHNISVHATEVHNRIHETRPVVAIKVLD 480
 gi | 1705549 | 421 IDRHITDLDRFFFITINPEDGFIKTKPLDREETAWLNISVFAAEIHNHQAETKVPVAIRVLD 480
 gi | 1377894 | 421 IDRHITDLDRFFFITINPEDGFIKTKPLDREETAWLNIITVFAAEIHNHQAERVPVAIRVLD 480

10

490 500 510 520 530 540
 NOV36a 472 BNDNAPQLAEFYDTFVCDSD---AAPGQLIQVIRALDRDEVGNSSHVSFQGP---LGPDA 524
 NOV36b 472 BNDNAPQLAEFYDTFVCDSD---AAPGQLIQVIRALDRDEVGNSSHVSFQGP---LGPDA 524
 gi | 16553903 | 467 -----W----- 468
 gi | 13626134 | 481 VNDNAPKFAAAVEAFVCEN---ARSNQFITISADKDDSANGPRFIFSLPEIITH-NP 535
 gi | 3377485 | 481 KNDNAPKFAKPYEAFVCEN---APINQEFITITAVDKDDTANGRLFLFSFPEIIVHPNP 536
 gi | 1705549 | 481 VNDNAPKFAAPYEGFITCESDHPKALSNOPIVTVSADDDDTANGPRFIFSLPEIIMH-NP 539
 gi | 1377894 | 481 VNDNAPKFAAPYEGFITCESDQTKPLSNQPIVTVSADDDDTANGPRFIFSLPEIITH-NP 539

15

20

550 560 570 580 590 600
 NOV36a 525 NFTVQDNRDGSASLLI-PSRPAPPRHAPYLVPIELWQWOPALSSSTATVTVSVCRCPQPDG 583
 NOV36b 525 NFTVQDNRDGSASLLI-PSRPAPPRHAPYLVPIELWQWOPALSSSTATVTVSVCRCPQPDG 583
 gi | 16553903 | 468 -----PRSCP-----WVWG----- 477
 gi | 13626134 | 536 NFSLRDNRDNTASVLVRREGFSRQKQDLYLLPIVTSDCGLPPMSSTNTLTIRVCGCDSDNG 595
 gi | 3377485 | 537 NFTIIDKRDNTASIRVGRGVFSRQKQDLYLLPIVTSDCGSPMSSTNTILSVRIKSCNSDG 596
 gi | 1705549 | 540 NFTVRDNRDNTAGVYARRGGFSRQKQDFYLLPIVTSDCGIPMSSTNTLTIRVCGCDVNG 599
 gi | 1377894 | 540 NFTVRDNRDNTAGVYARRGGFSRQKQDLYLLPIVTSDCGIPMSSTNTLTIRVCGCDVNG 599

25

30

610 620 630 640 650 660
 NOV36a 584 SVASCWPEAHLSAAGLSTGALLAITTCVGALLALVVLVVALRRQKQBALMVLEEDVREN 643
 NOV36b 584 SVASCWPEAHLSAAGLSTGALLAITTCVGALLALVVLVVALRRQKQBALMVLEEDVREN 643
 gi | 16553903 | 477 -----WRR-----VL----- 482
 gi | 13626134 | 596 SILSCNAEAYILNAGLSTGALLIATLACTIVILLVIVVLVFLRRQKKEPLIVFEEDVREN 655
 gi | 3377485 | 597 SOLSCNAEPQSLNAGLSTGALLIATLACTIVILLVIVVLVFLRRQKKEPLIVFEEDIREN 656
 gi | 1705549 | 600 ALLSCNAEAYILNAGLSTGALLIATLACTIVILLVIVVLVFLRRQKKEPLIVFEEDVREN 659
 gi | 1377894 | 600 ALLSCNAEAYILNAGLSTGALLIATLACTIVILLVIVVLVFLRRQKKEPLIVFEEDVREN 659

35

40

670 680 690 700 710 720
 NOV36a 644 IITYDDEGGGEEDTEAFDIATLQNPDGGAAPPAPGPPARRDVLPRARVSRQPRPGPADVA 703
 NOV36b 644 IITYDDEGGGEEDTEAFDIATLQNPDGGAAPPAPGPPARRDVLPRARVSRQPRPGPADVA 703
 gi | 16553903 | 482 --T-----AIAPSP----- 489
 gi | 13626134 | 656 IITYDDEGGGEEDTEAFDIATLQNPDGINGFIERKDIKPEYQYMPR-PGLRPAENSVDVD 714
 gi | 3377485 | 657 IITYDDEGGGEEDTEAFDIATLQNPDGINGFIERKDIKPEYQYMPR-DIGIRPAPNSVDVD 716
 gi | 1705549 | 660 IITYDDEGGGEEDTEAFDIATLQNPDGINGFIERKDIKPEYQYMPR-PGLRPAENSVDVD 718
 gi | 1377894 | 660 IITYDDEGGGEEDTEAFDIATLQNPDGINGFIERKDIKPEYQYMPR-PGLRPAENSVDVD 718

45

50

730 740 750 760 770 780
 NOV36a 704 QLIALRLREADEDPGVPPYDSVQVGYEGRGSSCGSLSSLSGSEAGGAPGPAEPLDDWG 763
 NOV36b 704 QLIALRLREADEDPGVPPYDSVQVGYEGRGSSCGSLSSLSGSEAGGAPGPAEPLDDWG 763
 gi | 16553903 | 489 -----ACGH----- 493
 gi | 13626134 | 715 DFINTRIQEADNDPTAPPYDSIQIYGYEGRGSVAGSLSSLESASTDSDLN--YDYLNQWG 772
 gi | 3377485 | 717 DFINTRIQEADNDPAAPPYDSIQIYGYEGRGSVAGSLSSLESASTDSDLN--YDYLNQWG 774
 gi | 1705549 | 719 DFINTRIQEADNDPTAPPYDSIQIYGYEGRGSVAGSLSSLESASTDSDLN--YDYLNQWG 776
 gi | 1377894 | 719 DFINTRIQEADNDPTAPPYDSIQIYGYEGRGSVAGSLSSLESASTDSDLN--YDYLNQWG 776

55

60

790 800
 NOV36a 764 PLFRTLAELYGAKEPPAP-- 781
 NOV36b 764 PLFRTLAELYGAKEPPAP-- 781
 gi | 16553903 | 493 ----- 493
 gi | 13626134 | 773 PRFKKLADLYGSKDTFDDDS 792
 gi | 3377485 | 775 PRFKKLADLYGSKDTCEDDS 794
 gi | 1705549 | 777 PRFKKLADLYGSKDTFDDDS 796
 gi | 1377894 | 777 PRFKKLADLYGSKDTFDDDS 796

65

Tables 36G-P list the domain description from DOMAIN analysis results against NOV36. This indicates that the NOV36 sequence has properties similar to those of other proteins known to contain this domain.

Table 36G. Domain Analysis of NOV36

gnl|Pfam|pfam01049, Cadherin_C_term, Cadherin cytoplasmic region.
Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn. (SEQ ID NO:827)
CD-Length = 150 residues, 98.7% aligned
Score = 99.8 bits (247), Expect = 5e-22

5	NOV36:	625	RRQKQEALMVLEEEDVRENIITYDDEGGGEEDTEAFDITALQNPDGAAPPAPGPPARRDV	684
			+ + ++ + + + + ++	
	Sbjct:	1	RRRKKEPLIIDEDDIRENIINYDDEGGGEEDTDAFDISALRSGGNPKPIEELKLRRDIK	60
10	NOV36:	685	LPRARVSRQPRPPGPADVAQLLALRLREADDPGVPPYDSVQVGYEGRGSSCGSLSSLG	744
			+ + + + + + +	
	Sbjct:	61	PELQSLPRPRRPAPDDIADFINEKLKEADNDPTAPPYDSLQTYAY--EGSGSVAGSLSS	118
15	NOV36:	745	SGSEAGGAPGPAEPLDDWGPLFRTLAELYG	774
			+ + + + ++	
	Sbjct:	119	LNSSTTSDQDYDYLNDWGPFRFKKLADMYG	148

Table 36H. Domain Analysis of NOV36

gnl|Pfam|pfam00028, cadherin, Cadherin domain (SEQ ID NO:828)
CD-Length = 92 residues, 97.8% aligned
Score = 81.3 bits (199), Expect = 2e-16

20	NOV36:	379	YHLTVPENKAPGTLVGQISAADLDSPA-SPIRYSILPHSDPERCFSIQPEEGTIHTAAPL	437
			+ ++ + +	
	Sbjct:	1	YSASVPENAPVGTEVLTVTATDADLGPNGRIFYSILGG-GPGGWFRIDPDTGDLSTTKPL	59
	NOV36:	438	DREARAWNLTVLATELGEDSQASRVQVAIQ	468
25			+ + + +	
	Sbjct:	60	DRESIGEYELTVLATDSGGPPLSGTTTIT	90

Table 36I. Domain Analysis of NOV36

gnl|Pfam|pfam00028, cadherin, Cadherin domain (SEQ ID NO:828)
CD-Length = 92 residues, 100.0% aligned
Score = 72.8 bits (177), Expect = 7e-14

30	NOV36:	264	YQFSVVETAGPGLVGRRLRAQDPDLGDNALMAYSILDGEGSEAFSISTDLQGRDGLLTVR	323
			+ + +	
	Sbjct:	1	YSASVPENAPVGTEVLTVTATDADLGPNGRIFYSILGGGPGGWFRIDPD----TGDLSTT	56
	NOV36:	324	KVLDFESQRSYSFRVEATNTLIDPAYLRRGPFKDVASVRVAVQ	366

77 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Table 36J. Domain Analysis of NOV36

gnl|Pfam|pfam00028, cadherin, Cadherin domain (SEQ ID NO:828)
CD-Length = 92 residues, 85.9% aligned
Score = 63.9 bits (154), Expect = 3e-11

10

Table 36K. Domain Analysis of NOV36

gnl|Pfam|pfam00028, cadherin, Cadherin domain (SEQ ID NO:828)
CD-Length = 92 residues, 97.8% aligned
Score = 59.3 bits (142), Expect = 8e-10

15

Table 36L. Domain Analysis of NOV36

gnl|Pfam|pfam00028, cadherin, Cadherin domain (SEQ ID NO:828)
CD-Length = 92 residues, 100.0% aligned
Score = 44.7 bits (104), Expect = 2e-05

25

Table 36M. Domain Analysis of NOV36

gnl|Smart|smart00112, CA, Cadherin repeats.; Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium. (SEQ ID NO:829)

CD-Length = 82 residues, 100.0% aligned
Score = 79.7 bits (195), Expect = 6e-16

| |

Sbjct: 57 TDGGG 61
Sbjct: 57 TDGGG 61

5 Cadherins, first discovered in mouse teratocarcinoma cells, are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion. Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. There are a number of different isoforms distributed in a tissue-specific manner in a wide variety of organisms. Cells containing different cadherins tend to segregate
10 in vitro, while those that contain the same cadherins tend to preferentially aggregate together. This observation is linked to the finding that cadherin expression causes morphological changes involving the positional segregation of cells into layers, suggesting they may play an important role in the sorting of different cell types during morphogenesis, histogenesis and regeneration. They may also be involved in the regulation of tight and gap junctions, and in
15 the control of intercellular spacing. Cadherins are evolutionary related to the desmogleins which are component of intercellular desmosome junctions involved in the interaction of plaque proteins. The first three cadherins to be described were *E-cadherin* is present on many types of epithelial cells; *N-cadherin* on nerve, muscle, and lens cells; and *P-cadherin* on cells in the placenta and epidermis.

20 The NOV36 proteins bear close resemblance to cadherin-11, a member of the cadherin family of proteins, expressed in osteoblasts. The tissue expression in brain, uterus and retina of these NOV36 proteins indicate they might play an important role during organogenesis and development.

The disclosed NOV36 nucleic acid of the invention encoding a cadherin 11-like
25 protein includes the nucleic acid whose sequence is provided in Table 36A, 36C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 36A or 36C while still encoding a protein that maintains its cadherin 11-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are
30 complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or
35 derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense

binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, in one embodiment up to about 35% of the NOV36a residues may be so changed and in an additional embodiment up to about 1% of the NOV36b residues may be so changed.

5 The disclosed NOV36 protein of the invention includes the cadherin 11-like protein whose sequence is provided in Table 36B or 36D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 36B or 36D while still encoding a protein that maintains its cadherin 11-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein,
10 up to about 43% of the NOV36a and NOV36b bases may be so changed.

 The above defined information for this invention suggests that these cadherin 11-like proteins (NOV36) is a member of a "cadherin 11 family". Therefore, the NOV36 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential
15 therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

20 The nucleic acids and proteins of NOV36 are useful in Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, endometriosis, fertility, anemia, ataxia-telangiectasia, autoimmune disease, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphoedema, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, and/or other pathologies and disorders.

25 NOV36 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV36 protein have multiple
30 hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV37

A disclosed NOV37 nucleic acid of 8575 nucleotides (also referred to as CG56733-01) encoding a novel Ten-M2-like protein is shown in Table 37A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 199-201 and ending with a TAA codon at nucleotides 8476-8478. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 37A, and the start and stop codons are in bold letters.

Table 37A. NOV37 Nucleotide Sequence (SEQ ID NO:143)

TAAAGTACCTGTCATCTTGACAAGTGGCGGAGCGGAGGAGTCAAGGATTATAAATGATCAGACCCAGGTCC
AGCTCGCCCCGTGATTGGGCTCTCCCGCGATCTGCACCGGGGAAGCGCATGAGAGGCCAATGAGACTTGA
ACCTGAGCCTAAGTTGTACCAGCAGGACTGATGTGCACACAGAGGAATGAAGTATGGATGTGAAGAA
CGCAGGCCTTACTGCTCCCTGACCAAGAGCAGACGAGAGAAGGAACGGCGCTACACAAATTCCTCCGCAGA
CAATGAGGAGTGCCGGGTACCCACACAGAAGTCCCTACAGTTCCAGCGAGACATTGAAAGCTTTTGATCATG
ATTCTCGCGGCTGCTTTACGGCAACAGAGTGAAGGATTGGTTCCAGAGAAAGCAGACGAGTTCACTAGA
CAAAGCAGGATGCACTATGGAACCGAGTCAAGACCTCATCCACCGGGAGTCAGATGAGTTTCTAGACA
AGGTATCCTTACCAGGGCTACTCCCTTAGCACAGGGTGTGACGCGACTCCGACACCGGAGGGAGGTGT
CTCCAGAACACGCCATCAGACTGTGGGGCAGAGGGATAAAATCCAGGCGCAGTTCGGGCTGTCCAGTCGT
GAAAACCTCGGCCCTTACCCTGACTGACTCTGACAACGAAACAAATCAGATGATGAGAACGGTCGTCCCAT
TCCACGTACATCCTCGCGTAGTCTCCTCCCATTGTTCAGCTGCCTAGCTCCCATAAATCTCCACAGTTA
GCTGCCAGATGCCATTGCTAGACAGCAACACCTCCCATCAAATCATGGACACCAACCTGATGAGGAATTC
TCCCCCAATTCATACCTGCTCAGAGCATGCTCAGGGCCCCAGCAAGCCTCCAGCAGTGGTCTCCGAACCA
CCACAGCCAGTCGACTCTGAGGCCCTCTCCACCCCTCACAACCAACGCTGTCCCATCACCACCTCGT
CCGCCAATCCCTCAACAGGAACCTCACTGACCAATCGGCGGAGTCAGATCCACGCCCGGCCAGCGCC
AATGACCTGGCCACCACACAGAGTCCGTTTCACTTTCAGGACAGCTGGGTGCTAAACAGCAACGTGCCACT
GGAGACCCGGCACTTCTCTTCAAGACCTCCTCGGGGAGCACACCCTTGTTCAGCAGCTCTTCCCGGGAT
ACCTTTGACCTCAGGAACGGTTTACACGCCCCCGCCCCGCTGTGCCAGGAATACTTTCTCCAGGAAG
GCTTTCAAGCTGAAGAAGCCCTCCAAATCTGACGTGGAATGTGCTGCCCTCTCCGCCATTGCCCGCGC
CCTGCTCTTGGCTATTTTGTGCTGGCTATTTTCGACGAATGCATCTGCTCGGACTCAATTGGCAACTCCAGC
CTGCAGATGGGCACACCTTTAACAATGGGATAAGGACCGGCTTACCAGGAACGATGATGTGGCAACAATG
CCATCTGGAGGCAAAGTGCCTGGTTCGTTGAAAAACAGCAGCATAGACAGTGGTGAAGCAGAAGTTGGTGC
CGGGTAACACAGAAGTCCACCAGGGGTGTTTGGAGGTCACAAATTCACATGCTCAGCCCCAGTCTCT
TAAAGTTCAACATCTCCCTCGGGAAGGACGCTCTCTTTGGTGTTTACATAAGAAGAGGACTTCCACCATCT
CATGCCAGTATGACTTTCATGGAACGCTCTGGACGGGAAGGAGAAGTGGAGTGTGTTGAGTCTCCAGGGA
ACGCGGGAGCATAACAGACCTTGGTTTCAAGATGAAGCCGTGTTTGTGCAGTACCTGGATGTGGGCTGTGGC
ATCTGCGCTTCTACAATGATGGAAGACAAAGAGATGGTTTCTTCAATACTGTTGTCTAGATTCACTG
CAGGACTGTCCACGTAACGCCATGGGAATGGTGAATGTGTGTCCGGGTGTGTCTACTGTTTCCAGGATT
TCTAGGAGCAGACTGTGCTAAAGCTGCCTGCCCTGTCTGTGTCAGTGGGAATGGACAATATTCTAAAGGGA
CGTGCCAGTGTCTACAGCGGCTGGAAGGTGCAGAGTGCAGCTGCCCATGAATCAGTGCATGCATCCTTCC
TGCGGGGGCCACGGCTCCTGCATTGATGGGAAGTGTGTCTGTCTGTCTGGCTACAAAGGCGAGCACTGTGA
GGAAGTTGATTGCTTGGATCCACCTGCTCCAGCCACGGAGTCTGTGTGAATGGAGAATGCCTGTGCAGCC
CTGGCTGGGGTGGTCTGAATGTGAGCTGGCGAGGGTCCAGTGCCAGACCAAGTGCAGTGGGCATGGCACG
TACCTGCCTGACACGGGCTCTGCAGTGCAGTCCCAACTGGATGGGTCCCGACTCTGTTGTGTGCTC
AGTAGACTGTGGCACTACGGCGTCTGCATCGGGGAGCCTGCCGCTGTGAAGAGGGCTGGACAGGCGCAG
CGTGTGACCAGCGCTGTGCCACCCCGCTGCATTGAGCACGGGACCTGTAAAGATGGCAAATGTGAATGC
CGAGAGGGCTGGAATGGTGAACACTGCACCATGATGGCTGCCCTGACTTGTGCAACGGTAACGGGAGATG
CACACTGGGTGAGAACAGCTGGCAGTGTGTCTGCCAGACCGGCTGGAGAGGGCCCGGATGCAACGTTGCCA
TGGAAACTTCTGTGCTGATAACAAGGATAATGAGGGAGATGGAATCATTGACTGCATGGATCCCGATTGC
TGCCTACAGAGTTCTGCCAGAATCAGCCCTATTGTCTGGGGAGTGCCTGATCCTCAGGACATCATTAGCCA
AAGCCTTCAATCGCTTCTCAGCAAGCTGCCAAATCCTTTTATGATCGAATCAGTTTCTTATAGGATCTG
ATAGACCCCATGTTATACCTGGAGAAAGTCTCTTCAATAGCTTGGTTTCTCTCATCCGAGGCCAAGTAGTA
ACTACAGATGGAACCTCCCTGGTGGTGTGAACGTGTCTTTTGTCAAGTACCCAAATACGGCTACACCAT
CACCCGCCAGGATGGCAGTACTCCCTCTCCAGGTTTCGACCTGATCGCAAATGGAGGTGCTTCTTGACTC
TACACTTTGAGCGAGCCCCGTTTCATGAGCCAGGAGCGCACTGTGTGGCTGCCGTGGACAGCTTTTACGCC
ATGGACCCCTGGTGTGATGAAGACCGAGGAGAACTCCATCCCCAGTTGTGACCTCAGTGGTTTTGTGCGCT
TGATCCCATCATCATCTCTCCCTCTGTCCACTTTCTTTAGTGTGCCCTGGGCAGAATCCCATCGTGC
CTGAGACCCAGGTAATCATGAAGAAATCGAGCTCCCTGGTTCGAATGTGAACTTCGCTATCTGAGCTCT
AGAATGCAGGGTACAAGTCACTGCTGAAGATCACCATGACCCAGTCCACAGTGCACCTGAACCTCATTAG
GGTTCACTGATGGTGGCTGTGAGGGGATCTCTTCCAGAAGTCAATCCAGGCTTCTCCCAACCTGGCCT
ACACCTTCATCTGGGACAAGACAGATGCGTATGGCCAAAGGGTGTATGGACTCTCAGATGCTGTTGGTATG
TTTTGGTTTCAAAGGACAGCCCTCCTTACAGGATTCGAGCTGGACCCCTCCAACCTCGGTGGCTGGTCCCT
AGACAAACACCACATCCTCAATGTTAAAGTGGTATCCTACACAAAGGCACTGGGGAAACAGTTCCTGA

CCACGACGCTGCCATCATACCAGCATCATATGGGCAATGGTCGCCGCCGGAGCATTTCCTGTCCACAGCTGC
AACGGCCCTTGCTGAAGGCAACAAGCTGCTGGCCCCAGTGGCTCTGGCTGTTTGAATCGATGGGAGCCTCTA
TGTGGGTGACTTTCATTTACATACAGCGCATCTTCCCTCTCGAAATGTGACCAAGCATCTTGGAGTTACGGA
GAAATAAAGACTTTTAAACATAGCAACAACCCAGCACAAGTACTTCTGGCAGTGGACCCCGTGTCCGGC
TCGCTCTACGTGTCCGACACCAACAGCAGGAGAATCTACCGCGTCAAGTCTCTGAGTGAACCAAAGACCT
GGCTGGGAATTCGGAAGTTTGTGGCAGGGACGGGAGAGCAGTGTCTACCCCTTGATGAAGCCCGCTGCGGGG
ATGGAGGAAGGCCATAGATGCAACCTGTATGAGCCCGAGAGGATTCAGTAGACAAGAATGGGCTCATG
TACTTTGTGATGCCACCATGATCCGGAAGGTTGACCAGAATGAATCATCTCCACCCTGCTGGGCTCCAA
TGACCTCACTGCCGTCCGGCCGCTGAGCTGTGATTCCAGCATGGATGTAGCCCAAGTTCGTCTGGAGTGGC
CAACAGACCTTGCTGTCAATCCCATGGATAACTCTTGATGTTCTAGAGAACAATGTCACTCTCGAATC
ACCGAAGAACCAAGTCAGCATATTGCGGGAGCGCCCATGCACTGCCAAGTTCCTGGCATTGACTACTC
ACTCAGCAAACTAGCCATTCACTCTGCCCTGGAGCTCAGCCAGTGCCATTGCCATTTCTACATGGGGTCC
TCTACATCACTGAGACAGATGAGAAGAAGATTAAACGCTCTACGCCAGGTAAACAACCAACGGGGAGATCTG
CTTTTAGCTGGGGCAGCTCCGGACTCGGACTCGCAATTCGCAAAACAGTGTCAATTTGCAACTGCTATTACGAGAGATGA
TGCCTACGCGACTGATGCCATCTTGAATCCCCATCACTTACTGTAGCTCCAGATGGGTACCAATTACA
TTGCAGACCTTGGAAATATTTCGGATCAGGGCGGTGACGAAGAACAAGCCTGTTCTTAATGCCCTTCAACAG
TATGAGGCTGCATCCCCCGGAGAGCAGGAGTTATATGTTTTCACGCTGATGGCATCCACCAATACACTGT
GAGCCTGGTGACAGGGGAGTACTTGTAACATTTACATATAGTACTGACAAATGATGTCTACTGAATTTGATT
ACAATAATGGGAATTCCTGAAGATCCGTGGGACGAGCAGTGGCATGCCCGTCACTTGTCTATGCTGTGAC
AACCAGATCATCACCTTCACTGTTGGGCACCAATGGAGGCCTCAAAGTCGTCTCCACACAGAACCTGGAGCT
TGGTCTATGACCTATGATGGCAACCATGGGCTCTGGCCACCAAGAGCGATGAACAGGATGGACGACTT
TCTATAGCTATGACCAGAAGGCCGCTGACCAACGTCAGCGCGCCCAACGGGGTGGTAACCACTCTGCAC
CGGAAATGGAGAATCTATTACCATTTGACATTTGAGAATCCAAACCGTGATGACGCTCACTGTCTATCAC
CAACCTCTCTTCAGTAGAGGCTCCTACACAGTGGTACAAGATCAAGTTCGGAACAGCTACCAGCTCTGTA
ATAATGTTACCTTGAGGGTGATGATGTCTAATGGTATGGGTATACAGTTCACACAGCGAGCCCCATGTCCTTA
CGGGGCACCATCACCCGCCACCATTTGACGCTGCAACATCTCCCTGCGCTATGGAAATGGCTTAACTCCAT
TGAGTGGCGCCTAAGAAAGGAACAGATTAAAGGCAAGTCACCATCTTTGGCAGGAAGCTCGAGGTCATG
GAAGAAATCTCTTGCTCAATTTGACTATGATCGAAATATTTCGGACTGAAAAGATCTATGATGACCACCGAAG
TTCACCTCTGAGGATCATTTATGACCAGGTGGGCGGCCCTCTCTCTGGCTGCCACAGCGGGCTGGCAGC
GTCAACGTGTCTACTTCTTCAATGGGCGCTGGCTGGGCTTCAGCGTGGGGGCCATGAGCGAGAGGACAG
ACATCGACAAGCAAGGCCGCATCGTGTCCCGCATGTTTCGCTGACGGGAAAGTGTGGAGCTACTCCTACCTT
GACAAGATGGTCTCTCTGCTTCAGAGCCAAAGTCAGTATATATTGAGTATGACTCTCTGACCGCCTCCT
TGCCCTCACCATGCCAGCGTGGCGCCGACAGCATGTCCACACACACCTCCATCGGCTACATCCGTATAA
TTTCAACCCCGCTGAAGAGCAATGCTTCGCTCATCTTTGACTACAGTGATGACGGCGCATCTCGAAGACC
TCCTTTTTTGGGCACCGGACGCCAGGTGTTCTACAAGTATGGGAAACTCTCCAAGTTATCAGAGATTGTCTA
CGCAGATACCGCGCTCACTTTCGGGTATGACGAGACCAGTGGTGTCTTGAAGATGGTCAACCTCCAAAGT
GGGGCTTCTCTGACCATTAGTACCGGAAGATTGGCCCCCTGGTGGAACAGCAGATCTACAGGTTCTCC
GAGGAAGGCATGGTCAATGCCAGGTTTGACTACACCTATCATGACAACAGCTTCGCATCGCAAGCATCAA
GCCCGTCATAAGTGAGACTCCCTTCCCGTTGACCTTACCGGTATGATGAGATTTCTGGCAAGGTGGAAC
ACTTTGGTAAAGTTTGGAGTCACTATTATGACATCAACAGATCATCACCATGCCGTGATGACCTTCAGC
AAACATCTGACACCACTAGGCGGATCAAGGAGTCCAGTATGAGATGTTCCGCTCCTCATGTATCTGAT
GACGGTGCAATATGACAGCATGGGCAGGGTGATCAAGAGGGAGCTAAAACCTGGGGCCTATGCCAATACCA
CGAAGTACACCTATGACTACGATGGGGACGGGCAGCTCCAGAGCGTGCCGGCGCTCAATGACCGCCGACC
TGGCGCTACAGCTATGACACTTAAATGGGAATCTCACTTACTGAAACCGGCAACAGTGTGCGCCTCATGCC
CTTGCGCTATGACCTTCGGGATCGGAATAACAGACTCGGGGATGTGAGTACAATAATGACGAGCATGGCT
ATCTGTGCCAGAGAGGGTCTGACATCTTCAATACAATTCCAAGGGCCTCCTAACAGAGCCTACAACAAG
GCCAGCGGGTGGAGTGTCCAGTACCGCTATGATGGCGTAGGACGGCGGGCTTCTTACAAGACAACCTGGG
CCACCACCTGCAGTACTTCTACTTGACTTCCACAACCCGACGCGCATCCCCATGCTACAACTCACTCCA
ACTCGGAGATTCACTCACTGTACTACGACCTTCAGGGCCACTCTTTGGCATGGAGAGCAGAGTGGGGAG
GAGTACTATGTTGCCTCTGATAACACAGGGACTCCTCTGGCTGTGTTTACGATCAACGGCCTCATGATCAA
ACAGTCTCAGTACACCGCTTATGGGAGATTATTAATGACTCCAACCCGACTTCCAGATGGTCAATTGGCT
TCCATTGGGGACTCTATGACCCCTGACCAAGCTGGTCCATCTCACTCAGCGTATTATGATGTGCTGGCA
GGACGATGGACCTCCCCAGACTATACCATGTGGGAAAAACGTGGGCAAGGAGCGGCCCCCTTTAACCTGTA
TATGTTCAAGAGCAACAATCTCTCAGCAGTGAGCTAGATTGAAGAACTACGTCAGACAGATGTGAAAAGCT
GGCTTGTGATGTTTGGATTTTCACTTTAGCAACATCTTCCGCTTCCGAGAGCAAAATGTATTTCTGTG
CCTCCTCCTATGAATTTGACAGAGTACAAGCAAGTGAGAATTGGACAGCTCATCACAGGTGCCACCAGAC
AACAGAGAGACATAAACAGGCTTTTATGGCTCTTGGAGGACAGGTCTATTACTAAAAAGCTCCACGCCAGCA
TCCGAGAGAAAGCAGGTCATTGGTTTGCCACCACCAACGCCCATATTGGCAAAGGCATCATGTTTGCCATC
AAGAAGGGCGGGTGACCAACGGGCGTGTCCAGATTCGCCAGCGAAGATGAGCCGCAAGGTGGCATCTGTGCT
GAACAACGGCTACTTACCTGGACAAGATGCATACAGCTCAGGGCGAAGGACACCCACTACTTTGTGAAGA
TTGGCTCAGCCGATGGCGACCTGGTCACACTAGGCACCACCATTGGGCCGCAAGGTGCTAGAGAGCGGGGTG
AACGTGACCGTGTCCCAGGCCACGCTGCTGGTCAACGGCAGGACTCGAAGGTTACGAAACATTGAGTTCCA
GTACTCCACGCTGCTGCTCAGCATCCGCTATGGCTCACCCTCCACCCGACCCCTGAGGAGAGAGGCCCCGG
TCTTGACACAGGCGAGACAGAGGGCCCTGGGCACGGCTGGGCCAAGGACGAGCAAGGACAGGACGGG
AGAGAGGGGAGCCGCTGTGGACTGAGGGCGAGAAGCAGCAGCTTCTGAGCACCGGGCGCGTGAAGGGTA
CGAGGATATTACGTGCTTCCCGTGGAGCAATACCAAGCTGTCAGACAGCTAGCAGCAACATCCAGTTT
TAAGACAGAATTGAGATGGGAAAGAGGTAACAAATAATCTGCTGCCATCTTGTCTGAATGGCTCAGCAG
GAGTAACCTGTTATCTCTCTTCTTAAGAGATGAGACCTAACAGGGGCACTGCGG

The NOV37 nucleic acid, located on chromosome 5, has 4965 of 5004 bases (99%) identical to a gb:GENBANK-ID:AB032953|acc:AB032953.1 mRNA from *Homo sapiens* (mRNA for KIAA1127 protein, partial cds) (E = 0.0).

A disclosed NOV37 polypeptide (SEQ ID NO:144) encoded by SEQ ID NO:143 is 2759 amino acid residues and is presented using the one-letter code in Table 37B. Signal P, Psort and/or Hydropathy results predict that NOV37 does not contain a signal peptide and is likely to be localized extracellularly with a certainty of 0.7900 in one embodiment, to the plasma membrane with a certainty of 0.7900 in another embodiment and to the nucleus with a certainty of 0.6000 in an additional embodiment.

Table 37B. Encoded NOV37 protein sequence (SEQ ID NO:144)

```
MDVKERRPYCSLTKSRREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHDSSRLLYGNRVKDLVHREAD
EFTQSRMHYGNRVTDLIHRESDEFPRQGIHQGYSLSTGSDADSDTEGGMSPHEAIRLWGRGIKSRSSGL
SSRENSALTTLTSDNENKSDDENGRPIPTSSRSLLPFVQLPSSHNPPPVSCQMPLLDSNTSHQIMDTNPDE
EFSNSYLLRACSGPQQASSSGPPNHHSQTLRPPLPPPHNHTLSHHSSANSNLNRSLTNRRSQIHAPAPA
PNDLATTPEVQLQDSWVLNSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGTVYTPPPRLLPRNTFSRK
AFKLKPKSKYCSWKCAALSAIAAALLAILLAYFAAMHLLGLNWQLQPADGHTFNNGIRTGLPGNDDVATMP
SGGKVPWSLKNSSIDSGEAEVGRRTQEVPPGVFWRFSQIHISQPPQFLKFNISLQKDALFGVYIRRGPLPSHA
QYDFMERLDGKEKWSVVEsprrrrsIQTLVQNEAVFVQYLDVGLWHLAFYNDGDKEMVSFNTVVLDSVQDC
PRNCHGNGECVSGVCHCFPGFLGADCAKACPVLCSGNGQYSGKTCQCYSGWGAECDVPMNQCIDPSCGGH
GSCIDGNCVCSAGYKGEHCEVDCLDPTCSSHGVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDT
GLCSCDPNWMGPDSCSVVCSVDCTHGVCIIGACRCEEGWTGAACDQRVCHPRCIEHGTCKDGKCECREGWNG
EHCTIDGCPDLCNGNGRCTLGQNSWQCVCQCTGWRGPGCNVAMETSCADNKDNEGDGLIDCMPDCCQLQSSCQ
NQPYCRGLPDPQDIISQSLQSPSQAAKSFYDRISFLIGSDSTHVI PGESPFNSLVSLIRGQVTTTDTGTPLV
GVNVSVFKYPKYGYTITRQDGTYSLSRFDLIANGGASLTLHFERAPFMSQERTVWLPWNSFYAMDTLVMKTE
ENSIPSCDLSGFCRLDPIIISPLSTFFSAAPGQNPVPEQVLEHEEIELPGSNVQLRYSRTVLSRLK
ITMTQSTVPLNLRVHLMVAVEGHLFQKSFQASPNLAYTFIWDKTDAYGQRVYGLSDAVGMFWFQRTALLQG
FELDPSNLGGWSLDKHHILNVKSGILHKGTEGENQFLTQQPAIITSIMGNRRRSISCPCSCNGLAEGNKLLAP
VALAVGIDGSLYVGDFNYIRRIFFSRNVTSILELRNKEFKHSNNPAHKYYLAVDPVSGSLYVSDTNSRRIY
RVKSLSGTKDLAGNSEVVAGTGEQCLPFDEARCGDGGKKAIDATLMSPRGIAVDKNGLMYFVDATMIRKVDON
GIISTLLGSNDLTAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYVLENNVILRITENHQVSIAGRPMH
CQVPGIDYLSKLAIHSALESASATAISHTGVLYITETDEKKINRLRQVTTNGEICLLAGAAASDCDCNKDVN
CNCYSGDDAYATDAILNSPSSLAAPDGTIYIADLGNIRIRAVSKNKPVLNAFNQYEAASPGEQELYFVND
GIHQYTVSLVTGEYLYNFTYSTDNVTELIDNNGNSLKRDSGMPRHLLMPDNQIITLTVGTNGGLKVVS
TQNLGLMTYDGNLTGLATKSDETGWTTFYSDHEGRLTNVTRPTGVVTSLHREMEKSIITIDIENSNRDDD
VTVITNLSSVEASYTVVDQVRNSYQLCNGGTLRVMYANGMGISFHSEPHVLGAGTITPTIGRCNISLPMENG
LNSIEWRLRKEQIKGKVTIFGRKLEVHGRNLLSIDYDRNIRTEKIYDHRKFTLRIIYDQVGRPFLWLPSG
LAAVNVSYFFNGRLAGLQRGAMSERTDIDKQGRIVSRMFADGKVWSYSYLDKMVLLLQSQRYIFEYDSSDR
LLAVTMPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRLKTSFLGTGRQVFYKYGKLSKLSEIV
YDSTAVTFGYDETTGVLKMNVLQSGGFSCITIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIK
PVISETPLPVDLYRYDEISGKVEHFGKFGVIYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLMYWMT
VQYDSMGRVIKRELKGPYANTTKYTYDYDGDGQLQSVPAVNDRTWRYSYDLNGLNHLNPGNSVRLMPLR
YDLRDRITRLGDVQYKIDDDGYLCQGRSDIFEYNSKGLLTRAYNKASGWSVQYRYDGVGRRASYKTNLGHHL
QYFYSDLHNPTRITHVYNHNSNEITSLYYDLQGHFLAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQY
TAYGEIYYDSNPDFQMVIGFHGGLYDPLTKLVHFTQRDYDVLAGRWTSPDYTMWKNVGKEPAPFNLYMFKSN
NPLSSELDLKNYVTDVKSVMFQGLSNIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHQTTERRHQ
AFMALEGQVITKKLHASIREKAGHWFATTTPIIGKIMFAIKEGRVTIGVSSIASEDSRKVASVLNNAYYLD
KMHYSIEGKDTHYFVKIGSADGDLVLTGTTIGRKVLESGVNVTVSQPTLLVNGRTRRFTNIEFOYSTLLLSI
RYGLTPDITLDEEKARVLDQARQALGTAWAKEQQKARDGREGSRLWTEGEKQQLSTGRVQGYEGYYVLPVE
QYPELADSSSNIQFLRQNMKGK
```

The disclosed NOV37 amino acid sequence has 2634 of 2764 amino acid residues (95%) identical to, and 2679 of 2764 amino acid residues (96%) similar to, the 2764 amino acid residue ptrn:SPTREMBL-ACC:Q9WTS5 protein from *Mus musculus* (Mouse) (TEN-M2) (E = 0.0).

NOV37 is predicted to be expressed in at least Amygdala, Brain, Bronchus, Cerebral Medulla/Cerebral white matter, Cochlea, Coronary Artery, Epidermis, Hair Follicles, Hippocampus, Hypothalamus, Kidney, Left cerebellum, Lung, Lymph node, Parietal Lobe, Pineal Gland, Retina, Right Cerebellum, Substantia Nigra, Vulva and Whole Organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, NOV37 is predicted to be expressed in the following tissues because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AB032953|acc: AB032953.1) a closely related *Homo sapiens* mRNA for KIAA1127 protein, partial cds homolog in species *Homo sapiens*: Amygdala, Brain, Bronchus, Cerebral Medulla/Cerebral white matter, Cochlea, Coronary Artery, Epidermis, Hair Follicles, Hippocampus, Hypothalamus, Kidney, Left cerebellum, Lung, Lymph node, Parietal Lobe, Pineal Gland, Retina, Right Cerebellum and Substantia Nigra.

NOV37 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 37C.

Table 37C. BLAST results for NOV37					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7657415 ref NP_035986.2 (NM_011856)	odd Oz/ten-m homolog 2 (Drosophila); odd Oz/ten-m homolog 3 (Drosophila) [<i>Mus musculus</i>]	2764	2633/277 7 (94%)	2677/2777 (95%)	0.0
gi 9910320 ref NP_064473.1 (NM_020088)	neurestin alpha [<i>Rattus norvegicus</i>]	2765	2625/277 8 (94%)	2676/2778 (95%)	0.0
gi 10241574 emb CAC09416.1 (AJ279031)	teneurin-2 [<i>Gallus gallus</i>]	2802	2525/281 8 (89%)	2639/2818 (93%)	0.0
gi 5307761 dbj BAA81892.1 (AB026979)	ten-m3 [Danio rerio]	2590	1707/260 2 (65%)	2097/2602 (79%)	0.0
gi 6760369 gb AAF28316.1 AF195418_1 (AF195418)	ODZ3 [<i>Mus musculus</i>]	2346	1665/236 0 (70%)	1993/2360 (83%)	0.0

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 37D.

Table 37D Clustal W Sequence Alignment

1) NOV37 (SEQ ID NO:144)
2) gi|7657415|ref|NP_035986.2| (NM_011856) odd Oz/ten-m homolog 2 (Drosophila); odd Oz/ten-m homolog 3 (Drosophila) [*Mus musculus*] (SEQ ID NO:487)
3) gi|9910320|ref|NP_064473.1| (NM_020088) neurestin alpha [*Rattus norvegicus*] (SEQ ID NO:488)

5

			10	20	30	40	50	60																																																			
10	NOV37	1	MDV	KERR	PYCS	LT	KSR	REK	ERRY	TN	SSA	DN	EE	CR	VP	TQ	KSY	SS	SET	LK	AF	DH	DS	SR	LL	YC	60																																
	gi 7657415	1	MDV	KDRR	-HRS	LT	RGR	CG	KE	CRY	TSS	SL	DS	ED	CR	VP	TQ	KSY	SS	SET	LK	AY	DH	DS	-RM	HY	58																																
	gi 9910320	1	MDV	KDRR	-HRS	LT	RGR	CG	KE	CRY	TSS	SL	DS	ED	CR	VP	TQ	KSY	SS	SET	LK	AY	DH	DS	-RM	HY	58																																
	gi 10241574	1	MDV	KDRR	-HRS	LT	RGR	CG	KE	CRY	TSS	SL	DS	ED	CR	VP	AQ	KSY	SS	SET	LK	AY	GHD	-RM	HY	58																																	
	gi 5307761	1	-----																																																								
	gi 6760369	1	-----																																																								
15			-----																																																								
			70	80	90	100	110	120																																																			
20	NOV37	61	NRV	KDL	VH	RE	AD	EF	TR	Q	-SR	M	HY	G	N	R	-V	T	D	L	I	H	R	E	S	D	E	F	PR	Q	G	I	L	H	Q	G	Y	S	L	T	G	S	D	A	D	S	118												
	gi 7657415	59	NRV	T	D	L	V	H	R	E	S	D	E	F	S	R	Q	G	T	N	F	T	L	A	E	L	G	I	C	E	P	S	P	H	R	S	G	Y	C	S	D	M	G	I	L	H	Q	G	Y	S	L	T	G	S	D	A	D	S	118
	gi 9910320	59	NRV	T	D	L	V	H	R	E	S	D	E	F	S	R	Q	G	A	N	F	T	L	A	E	L	G	I	C	E	P	S	P	H	R	S	G	Y	C	S	D	M	G	I	L	H	Q	G	Y	S	L	T	G	S	D	A	D	S	118
	gi 10241574	59	NRV	S	D	L	V	H	R	E	S	D	E	F	S	R	Q	G	T	N	F	T	L	A	E	L	G	I	C	E	P	S	P	H	R	S	G	Y	C	S	D	I	G	I	L	H	Q	G	Y	S	L	T	G	S	D	A	D	S	118
	gi 5307761	1	-----																																																								
	gi 6760369	1	-----																																																								
25			-----																																																								
			130	140	150	160	170	180																																																			
30	NOV37	119	T	E	G	G	M	S	P	E	H	A	I	R	L	W	G	R	G	I	K	S	R	S	S	G	L	S	S	R	E	N	S	A	L	T	L	T	D	S	N	E	N	K	S	D	D	N	-----	167									
	gi 7657415	119	T	E	G	G	M	S	P	E	H	A	I	R	L	W	G	R	G	I	K	S	R	S	S	G	L	S	S	R	E	N	S	A	L	T	L	T	D	S	N	E	N	K	S	D	D	N	-----	167									
	gi 9910320	119	T	E	G	G	M	S	P	E	H	A	I	R	L	W	G	R	G	I	K	S	R	S	S	G	L	S	S	R	E	N	S	A	L	T	L	T	D	S	N	E	N	K	S	D	D	N	-----	167									
	gi 10241574	119	T	E	G	G	M	S	P	E	H	A	I	R	L	W	G	R	G	I	K	S	R	S	S	G	L	S	S	R	E	N	S	A	L	T	L	T	D	S	N	E	N	K	S	D	E	E	N	D	F	H	T	L	S	E	K	L	178
	gi 5307761	1	-----																																																								
	gi 6760369	1	-----																																																								
35			-----																																																								
			190	200	210	220	230	240																																																			
40	NOV37	167	-----																																																								
	gi 7657415	167	-----																																																								
	gi 9910320	167	-----																																																								
	gi 10241574	179	D	R	Q	T	S	W	Q	L	A	E	T	K	N	S	L	I	R	P	I	P	T	S	S	S	L	L	P	S	A	Q	L																										

5	NOV37	389	ILLAYFAAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATMPSSGK---VPWSLKNS	445
	gi 7657415	389	ILLAYFIAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATVPSSGK---VPWSLKNS	445
	gi 9910320	389	ILLAYFIAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATVPSSGK---VPWSLKNS	445
	gi 10241574	417	ILLAYFIAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATVPSSGK---VPWSLKNS	473
	gi 5307761	216	ILLAYFIAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATVPSSGK---VPWSLKNS	272
10	gi 6760369	1	ILLAYFIAMHLLGLNWQLPADGHTFNNCIRTGLPCNDDVATVPSSGK---VPWSLKNS	29
			490 500 510 520 530 540	
	NOV37	446	IDSGEAEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	505
	gi 7657415	446	IDSGEAEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	505
	gi 9910320	446	IDSGEAEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	505
15	gi 10241574	474	IDSGETEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	533
	gi 5307761	273	IDSGETEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	332
	gi 6760369	30	IDSGETEVGRRTQEVPPGVFWRSQIHISQPFKFNISLGKDALFGVYIRRGLPSSHAQ	89
			550 560 570 580 590 600	
	NOV37	506	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	560
20	gi 7657415	506	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	560
	gi 9910320	506	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	560
	gi 10241574	534	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	588
	gi 5307761	333	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	392
25	gi 6760369	90	YDFMERLDG-----KEKWSVVEPRRERSIQTLVQNEAVFVQYLDVGLWHLAFYNDGKDK	149
			610 620 630 640 650 660	
	NOV37	561	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	620
	gi 7657415	561	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	620
	gi 9910320	561	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	620
30	gi 10241574	589	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	648
	gi 5307761	393	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	452
	gi 6760369	150	EMVSFNTVVLDSVQDCPRNCHNGECVSGVCHCFPGFLGADCAKAACPVLCSGNGQYSKG	209
			670 680 690 700 710 720	
	NOV37	621	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	680
40	gi 7657415	621	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	680
	gi 9910320	621	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	680
	gi 10241574	649	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	708
	gi 5307761	453	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	512
45	gi 6760369	210	TCOCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSAGYKGEHCEEVDCLDPTCSSH	269
			730 740 750 760 770 780	
	NOV37	681	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	739
	gi 7657415	681	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	739
	gi 9910320	681	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	740
50	gi 10241574	709	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	768
	gi 5307761	513	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	572
	gi 6760369	270	GVCVNGECLCSPGWGGLNCELARVQCPDQCSGHGTYLPDGLCSGCDPNWMMGPDCSV-VCS	329
			790 800 810 820 830 840	
	NOV37	740	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	797
55	gi 7657415	740	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	797
	gi 9910320	741	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	798
	gi 10241574	769	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	828
	gi 5307761	573	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	630
60	gi 6760369	330	VDCGTHGVCIGGACRCEEGWTGAACDQVCHPRCTEHGTCKDGKCECREGWNGEHCTI	387
			850 860 870 880 890 900	
	NOV37	797	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	850
	gi 7657415	797	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	850
	gi 9910320	798	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	851
65	gi 10241574	829	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	888
	gi 5307761	630	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	683
	gi 6760369	387	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	440
			850 860 870 880 890 900	
	NOV37	797	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	850
70	gi 7657415	797	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	850
	gi 9910320	798	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	851
	gi 10241574	829	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	888
	gi 5307761	630	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	683
	gi 6760369	387	DGCPDLNCGNGRCTLGQNSWQVCVQGWGPGCNVAMETSCADNKNDEGDGLV	440

			910	920	930	940	950	960		
5	NOV37	851	DCMDPDCCLQSSCONQPYCRGLPDPQDIIS	SLQSPSQAAKSFYDRISFLIGSDSTHVI	910					
	gi 7657415	851	DCLDPDCCLQSAQCONSLLCRGSRDPLDII	IQGQ--TDWPAVKSFYDRIKLLAGKSTHII	908					
	gi 9910320	852	DCLDPDCCLQSAQCONSLLCRGSRDPLDII	IQGQ--TDWPAVKSFYDRIKLLAGKSTHII	909					
	gi 10241574	889	DCLVPDCCLQSTCONSLLCRGSRDPLDII	IQSH--SGSPA VKSFYDRIKLLVCKDSTHII	946					
	gi 5307761	684	DCMDPDCCLQSSCQTPFCRGSPPDFDIIS	QNPAS PQAAQSFYQGISFLITGPSTHVI	743					
10	gi 6760369	441	DCMDPDCCLQSSCONQPYCRGLPDPQDIIS	SLQTPSQAAKSFYDRISFLIGSDSTHVI	500					
			970	980	990	1000	1010	1020		
15	NOV37	911	PGESPPFN	SLVSLIRGQVVT	DGTPLVGVNVSVFKYPKYGYTITRQDGT	TYSLSR	FDLI	AN	969	
	gi 7657415	909	PGDNPFNS	SLVSLIRGQVVT	DGTPLVGVNVSVFKYPKYGYTITRQDGT	-----	FDLI	AN	963	
	gi 9910320	910	PGDNPFNS	SLVSLIRGQVVT	DGTPLVGVNVSVFKYPKYGYTITRQDGT	-----	FDLI	AN	964	
	gi 10241574	947	PGENPFNS	SLVSLIRGQVVT	DGTPLVGVNVSVFKYPKYGYTITRQDGM	-----	FDLI	AN	1001	
	gi 5307761	744	NGENPFNR	SLVSIIRGQVLT	ADGTPLIGVNVSVFHYDPHCYTTITRQDGM	-----	FDLI	AN	798	
	gi 6760369	501	PGESPPFN	KLASVIRGQVLT	ADGTPLIGVNVSVFHYSEYGYTITRQDGM	-----	FDLI	AN	555	
20			1030	1040	1050	1060	1070	1080		
	NOV37	970	GGASLTTLHFERAPFMSQERTVWLPWNSFYAMDTLVMKTEENSIPSCDLSGFCRLDPIIIS	1029						
	gi 7657415	964	GGASLTTLHFERAPFMSQERTVWLPWNSFYAMDTLVMKTEENSIPSCDLSGFCRLDPIIIS	1023						
	gi 9910320	965	GGASLTTLHFERAPFMSQERTVWLPWNSFYAMDTLVMKTEENSIPSCDLSGFCRLDPIIIS	1024						
25	gi 10241574	1002	GGASLTTLHFERAPFMSQERTVWLPWNSFYAMDTLVMKTEENSIPSCDLSGFCRLDPIIIS	1061						
	gi 5307761	799	GGASLTLSFERAPFLTQFRVWLPWNVFYVMDTLVMKTEENDIPSCDLSGFIRESPLIIV	858						
	gi 6760369	556	GGASLTTLVFERSPFLTQYHTVWLPWNVFYVMDTLVMKTEENDIPSCDLSGFCRLDPIIIS	615						
30			1090	1100	1110	1120	1130	1140		
	NOV37	1030	SPLSTFFS	AAPGONPIVPETQVLHHEEIELPGSNV	KLYLSSRTAGYKSLKTI	MTQSTVP	1089			
	gi 7657415	1024	SPLSTFFS	ASPAENPIVPETQVLHHEEIELPGTNV	KLYLSSRTAGYKSLKTI	MTQSTVP	1083			
	gi 9910320	1025	SPLSTFFS	ASPAENPIVPETQVLHHEEIELPGTNV	KLYLSSRTAGYKSLKTI	MTQSTVP	1084			
	gi 10241574	1062	SPLSTFFS	DAPGRNPIVPETQVLHHEEIEVPGSSI	KLYLSSRTAGYKSLKTI	MTQSLVP	1121			
35	gi 5307761	859	TPLSTFFRS	SPENGPITPETQVLQEEETAI	PGSDLNMYLSSRAAGYR	PVLKVTMTQATIP	918			
	gi 6760369	616	SPLSTFFRS	SPEDSPITPETQVLHEETTIPGTDL	KLSYLSSRAAGYKSLKTI	MTQAVIP	675			
40			1150	1160	1170	1180	1190	1200		
	NOV37	1090	LNLIRVHLMVAVEGHLFQKSFQASP	NLAYTFIWDKTDAYGQ	RVYGLSDAVG	-----	1140			
	gi 7657415	1084	LNLIRVHLMVAVEGHLFQKSFQASP	NLAYTFIWDKTDAYGQ	RVYGLSDAVVSGFEYETC	1143				
	gi 9910320	1085	LNLIRVHLMVAVEGHLFQKSFQASP	NLAYTFIWDKTDAYGQ	RVYGLSDAVVSGFEYETC	1144				
	gi 10241574	1122	LNLIKVHLMVAVEGHLFQKSFQASP	NLAYTFIWDKTDAYGQ	KVYGLSDAVVSGFEYETC	1181				
	gi 5307761	919	FNLMKVHLMVAVVGRIFQKWFPA	PNLSYTFIWDKTDAYN	QVYGLSEAVVSGFEYESC	978				
45	gi 6760369	676	FNLMKVHLMVAVVGRIFQKWFPA	PNLAYTFIWDKTDAYN	QVYGLSEAVVSGFEYESC	735				
50			1210	1220	1230	1240	1250	1260		
	NOV37	1140	---MFWFQRTALLOGFELDP	SNLGGWSLDKHHTLN	NVKS	GILHKG	TGENQFLTQQPA	IITS	1197	
	gi 7657415	1144	PSLILWEKRTALLOGFELDP	SNLGGWSLDKHHTLN	NVKS	GILHKG	TGENQFLTQQPA	IITS	1203	
	gi 9910320	1145	PSLILWEKRTALLOGFELDP	SNLGGWSLDKHHTLN	NVKS	GILHKG	TGENQFLTQQPA	IITS	1204	
	gi 10241574	1182	PSLILWEKRTALLOGFELDP	SNLGGWSLDKHHTLN	NVKS	GILHKG	TGENQFLTQQPA	IITS	1241	
	gi 5307761	979	LDLILWEKRTALLOGFELD	ASNMGGWILDKHHVLD	VQNGILYKNGEN	VFVSQOP	PVIST	1038		
55	gi 6760369	736	LDLILWEKRTAVLOGFELD	ASNMGGWILDKHHVLD	VQNGILYKNGEN	QFISQOP	PVVSS	795		
60			1270	1280	1290	1300	1310	1320		
	NOV37	1198	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALAVGIDGSLV	VGDFNYIRRI	FPSRNV	TSILE	1257	
	gi 7657415	1204	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALAVGIDGSLV	VGDFNYIRRI	FPSRNV	TSILE	1263	
	gi 9910320	1205	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALAVGIDGSLV	VGDFNYIRRI	FPSRNV	TSILE	1264	
	gi 10241574	1242	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALAVGIDGSLV	VGDFNYIRRI	FPSRNV	TSILE	1301	
	gi 5307761	1039	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALACGSDGSLV	VGDFNYIRRI	FPSRNV	TSILE	1098	
	gi 6760369	796	IMGNGRRRS	ISCPSCNGLAEGNKLLAP	VALACGSDGSLV	VGDFNYIRRI	FPSRNV	TSILE	855	
65			1330	1340	1350	1360	1370	1380		
	NOV37	1258	LRRNKEFKHSNN	PAHKYYLAVDPV	SGSLYVSDTNS	RRIRYVKSL	SGTKDL	LAGNSEV	VAGT	1317
	gi 7657415	1264	LR-NKEFKHSN	SPGHKYYLAVDPV	TGSLYVSDTNS	RRIRYVKSL	SGAKDL	LAGNSEV	VAGT	1322
	gi 9910320	1265	LR-NKEFKHSN	SPGHKYYLAVDPV	TGSLYVSDTNS	RRIRYVKSL	SGAKDL	LAGNSEV	VAGT	1323
70	gi 10241574	1302	LR-NKEFKHSN	PAHKYYLAVDPV	TGSLYVSDTNS	RRIRYVKSL	SGTKDL	LAGNSEV	VAGT	1360

		gi 5307761	1099	LS-----NN-PAHGYLLATDPMTGOLYVSDTNSRRIRPKALITGKELQNAEUVAGT	1150	
		gi 6760369	856	LS-----SN-PAHRYLLATDPVTGDLVSDTNRRIYRPKSLTGAKDLTKNAEUVAGT	907	
		<div> <div>139014001410142014301440</div> <div>..... </div> </div>				
5	NOV37		1318	GEQCLPFDEARCGDGGKAVDATALMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGS	1377	
	gi 7657415		1323	GEQCLPFDEARCGDGGKAVDATALMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGS	1382	
	gi 9910320		1324	GEQCLPFDEARCGDGGKAVDATALMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGS	1383	
	gi 10241574		1361	GEQCLPFDEARCGDGGKAVDATALMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGS	1420	
10	gi 5307761		1151	GEQCLPFDEARCGDGGKATEALLGPKGIAVDKNGFTYFVDGTMIRKVDQNGIISTLLGS	1210	
	gi 6760369		908	GEQCLPFDEARCGDGGKAVEATLMSPKGMAIDKNGLYFVDGTMIRKVDQNGIISTLLGS	967	
		<div> <div>145014601470148014901500</div> <div>..... </div> </div>				
15	NOV37		1378	NDLTAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYLVLENNVILRITENHQVSI IAGR	1437	
	gi 7657415		1383	NDLTAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYLVLENNVILRITENHQVSI IAGR	1442	
	gi 9910320		1384	NDLTAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYLVLENNVILRITENHQVSI IAGR	1443	
	gi 10241574		1421	NDLTAVRPLSCDSSMDVSVQRLEWPTDLAVNPMDNSLYLVLENNVILRITENHQVSI IAGR	1480	
	gi 5307761		1211	NDLTSARPLTCDNSMHTGQVRLEWPTDLAINPMDNSLYVLNNNVILQITENRQVRI IAGR	1270	
20	gi 6760369		968	NDLTSARPLTCDNSMHTGQVRLEWPTDLAINPMDNSLYVLNNNVILQITENRQVRI IAGR	1027	
		<div> <div>151015201530154015501560</div> <div>..... </div> </div>				
25	NOV37		1438	PMHCQVPGIDYSLSKLAIHSALESASAIASHTGCVLYITETDEKKINRLRQVTTNGEICL	1497	
	gi 7657415		1443	PMHCQVPGIDYSLSKLAIHSALESASAIASHTGCVLYITETDEKKINRLRQVTTNGEICL	1502	
	gi 9910320		1444	PMHCQVPGIDYSLSKLAIHSALESASAIASHTGCVLYITETDEKKINRLRQVTTNGEICL	1503	
	gi 10241574		1481	PMHCQVPGIDYSLSKLAIHSALESASAIASHTGCVLYITETDEKKINRLRQVTTNGEICL	1540	
	gi 5307761		1271	PMHCQVPGIETVMGKRAIQTTLGATAIISLSYSGVLYIAETDEKKINRLRQVTTNGEISH	1330	
30	gi 6760369		1028	PMHCQVPGVEYFVCKHAWQTTLESATAIAVSYSGVLYITETDEKKINRLRQVTTNGEIST	1087	
		<div> <div>157015801590160016101620</div> <div>..... </div> </div>				
35	NOV37		1498	LAGAASDCDCCKNDVNCI CYSGDDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSK	1557	
	gi 7657415		1503	LAGAASDCDCCKNDVNCI CYSGDDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSK	1562	
	gi 9910320		1504	LAGAASDCDCCKNDVNCI CYSGDDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSK	1563	
	gi 10241574		1541	LAGAASDCDCCKNDVNCI CYSGDDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSK	1600	
	gi 5307761		1331	LAGAASDCDCCKNDANDCYOTGDGAKDARLNAPSSSLVSPDGTIYIADLGNIRIRAIRH	1390	
	gi 6760369		1088	VAGIPSECDCKNDANDCYQSGDGYAKDAKLNAPSSSLAASPDGTIYIADLGNIRIRAVSK	1147	
		<div> <div>163016401650166016701680</div> <div>..... </div> </div>				
40	NOV37		1558	NKPLVILNAFNOYEAAASPGQEQLYVFNADGIHQYTVSLVTGGEYLYNFYTSNDNDVTELDINN	1617	
	gi 7657415		1563	NKPLVILNAFNOYEAAASPGQEQLYVFNADGIHQYTVSLVTGGEYLYNFYTSADNDVTELDINN	1622	
	gi 9910320		1564	NKPLVILNAFNOYEAAASPGQEQLYVFNADGIHQYTVSLVTGGEYLYNFYTSADNDVTELDINN	1623	
45	gi 10241574		1601	NRPLILNSFNOYEAAASPGQEQLYVFNADGIHQYTVSLVTGGEYLYNFYTSNDNDVTEVMDSN	1660	
	gi 5307761		1391	NRPPQGSGLLEVASPASQEQLYVFDSNGTHQYTMVSLVTGGEYLYNFYTSNEDDVTAITDSS	1450	
	gi 6760369		1148	NKPLILNSMNFYEVASPTDQEQLYIFDINGTHQYTVSLVTGGEYLYNFYTSNDNDVTAITDSS	1207	
		<div> <div>169017001710172017301740</div> <div>..... </div> </div>				
50	NOV37		1618	GNSLKIIRRDSSGMPRHLLMPDNQIITLTGVTNGGLKAVSTQNLGLMTYDGNITGLLATK	1677	
	gi 7657415		1623	GNSLKIIRRDSSGMPRHLLMPDNQIITLTGVTNGGLKAVSTQNLGLMTYDGNITGLLATK	1682	
	gi 9910320		1624	GNSLKIIRRDSSGMPRHLLMPDNQIITLTGVTNGGLKAVSTQNLGLMTYDGNITGLLATK	1683	
	gi 10241574		1661	GNSLKIIRRDSSGMPRHLLMPDNQIITLTGVTNGGLKAVSTQNLGLMTYDGNITGLLATK	1720	
55	gi 5307761		1451	GNTLRIIRRDPPRMFVRVSPDNQVIWLTIGTNGCLKSMIAQGLLEVLFTYHGNISGLLATK	1510	
	gi 6760369					

	gi 9910320	1744	VEASYTVVDQVRNSYQLCSNGTLRVMYANGMGVSFHSEPHVLAGTLTPTTIGRCNISLPM	1803			
	gi 10241574	1781	VEASYTVVDQVRNSYQLCSNGTLRVMYANGMGVSFHSEPHVLAGTLTPTTIGRCNISLPM	1840			
	gi 5307761	1571	IDSFYTLVQDQLRNSYQVGYDMSRVIYANGMDSHFQTEPHILACASNPTVARRNMTLPG	1630			
5	gi 6760369	1328	IDSFYTMVQDQLRNSYQIGYDGLRIFVASCLEDSHVOTEPHVLAGTANPTVARRNMTLPG	1387			
		1870	1880	1890	1900	1910	1920
	NOV37	1798	ENGLNSIEWRLRKEQIKGKVTFGRKLEVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIT	1857			
10	gi 7657415	1803	ENGLNSIEWRLRKEQIKGKVTFGRKLRVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIT	1862			
	gi 9910320	1804	ENGLNSIEWRLRKEQIKGKVTFGRKLRVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIT	1863			
	gi 10241574	1841	ENGLNSIEWRLRKEQIKGKVTFGRKLRVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIT	1900			
	gi 5307761	1631	ENGONLVEWRFRKEQNRGKVNVFGRKLRVNGRNLLSIDYDRSLRTEKIYDDHRKFTLRIT	1690			
	gi 6760369	1388	ENGONLVEWRFRKEQAQGVNVFGRKLRVNGRNLLSIDYDRSLRTEKIYDDHRKFTLRIT	1447			
		1930	1940	1950	1960	1970	1980
	NOV37	1858	YDQVGRPFLWLPSSGLAAVNVSFFNGRLAGLQAGMSERTDIDKQGRIVSRMFADGKVV	1917			
	gi 7657415	1863	YDQVGRPFLWLPSSGLAAVNVSFFNGRLAGLQAGMSERTDIDKQGRIVSRMFADGKVV	1922			
20	gi 9910320	1864	YDQVGRPFLWLPSSGLAAVNVSFFNGRLAGLQAGMSERTDIDKQGRIVSRMFADGKVV	1923			
	gi 10241574	1901	YDQVGRPFLWLPSSGLAAVNVSFFNGRLAGLQAGMSERTDIDKQGRIVSRMFADGKVV	1960			
	gi 5307761	1691	YDASCHPTLWLPSSKLMVNVITSSSTGQVTSLQRCPTTTERVEYDSQGRIVSRMFADGKVV	1750			
	gi 6760369	1448	YDTSCHPTLWLPSSKLMVNVITSSSTGQVTSLQRCPTTTERVEYDSQGRIVSRMFADGKVV	1507			
		1990	2000	2010	2020	2030	2040
25	NOV37	1918	SYSYLDK-MVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	1976			
	gi 7657415	1923	SYSYLDKSMVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	1982			
	gi 9910320	1924	SYSYLDKSMVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	1983			
	gi 10241574	1961	SYTYLEKSMVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	2020			
30	gi 5307761	1751	SYTYLDR-MVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	1810			
	gi 6760369	1508	SYTYLEKSMVLLLSQSRQYIFEYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPES	1567			
		2050	2060	2070	2080	2090	2100
35	NOV37	1977	NASVIFDYSDGGRILKTSFLGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLKVV	2036			
	gi 7657415	1983	NASVIFDYSDGGRILKTSFLGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLKVV	2042			
	gi 9910320	1984	NASVIFDYSDGGRILKTSFLGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLKVV	2043			
	gi 10241574	2021	NASVIFDYSDGGRILKTSFLGTGRQVFKYKGLSKLSEIYVDSTAVTFGYDETTGVLKVV	2080			
	gi 5307761	1811	NASVITVDYSEDGQLLRVAHLGTGRVLYKYRQNKLSLSEIYVDSTAVTFGYDETTGVLKVV	1870			
40	gi 6760369	1568	NASVITVDYSEDGQLLRVAHLGTGRVLYKYRQNKLSLSEIYVDSTAVTFGYDETTGVLKVV	1627			
		2110	2120	2130	2140	2150	2160
45	NOV37	2037	NLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	2096			
	gi 7657415	2043	NLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	2102			
	gi 9910320	2044	NLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	2103			
	gi 10241574	2081	NLQSGGFSCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	2140			
	gi 5307761	1871	NLQSEGFICTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	1929			
	gi 6760369	1628	NLQSEGFICTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRIASIKPVISETPL	1686			
50		2170	2180	2190	2200	2210	2220
	NOV37	2097	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	2156			
	gi 7657415	2103	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	2162			
55	gi 9910320	2104	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	2163			
	gi 10241574	2141	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	2200			
	gi 5307761	1930	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	1989			
	gi 6760369	1687	PVDLYRYDEISGKVEHFGKFGVIYYDINQIITAVMTLSKHFDTHGRIKEVQYEMFRSLM	1746			
		2230	2240	2250	2260	2270	2280
60	NOV37	2157	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	2216			
	gi 7657415	2163	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	2221			
	gi 9910320	2164	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	2222			
65	gi 10241574	2201	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	2259			
	gi 5307761	1990	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	2048			
	gi 6760369	1747	YWMTVQYDSMGRVTKRELKLGPIYANTTKYTYDYDGDGQLQSV-PAVNDRPTWRYSYDLNGN	1805			
		2290	2300	2310	2320	2330	2340
70							

5	NOV37	2217	LHLLNPGNSVRLMPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	2276
	gi 7657415	2222	LHLLNPGNSARLMPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	2281
	gi 9910320	2223	LHLLNPGNSARLMPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	2282
	gi 10241574	2260	LHLLNPGNSVRLMPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	2319
	gi 5307761	2049	LHLLNPGNSARLTPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	2108
10	gi 6760369	1806	LHLLNPGSSARLTPLRYDLRDRITRLGDOYKIDDDGYLCQSGSDIFEYNSKGLLTRAYN	1865
		 2350 2360 2370 2380 2390 2400	
	NOV37	2277	KASGWSVOYRYDGVGRRASYKTNLGHHLQYFYSDLHNPTRITHVYNHNSSEITSLYYDLQ	2336
	gi 7657415	2282	KASGWSVOYRYDGVGRRASYKTNLGHHLQYFYSDLHNPTRITHVYNHNSSEITSLYYDLQ	2341
	gi 9910320	2283	KASGWSVOYRYDGVGRRASYKTNLGHHLQYFYSDLHNPTRITHVYNHNSSEITSLYYDLQ	2342
15	gi 10241574	2320	KANGWNVOYRYDGLGRRASCKTNLGHHLQYFYADLHNPTRITHVYNHNSSEITSLYYDLQ	2379
	gi 5307761	2109	KASGWTVOYRYDGLGRRASCKTNLGHHLQYFYADLHNPTRITHVYNHNSSEITSLYYDLQ	2168
	gi 6760369	1866	KASGWTVOYRYDGLGRRASCKTNLGHHLQYFYADLHNPTRITHVYNHNSSEITSLYYDLQ	1925
		 2410 2420 2430 2440 2450 2460	
	NOV37	2337	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	2396
20	gi 7657415	2342	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	2401
	gi 9910320	2343	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	2402
	gi 10241574	2380	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	2439
	gi 5307761	2169	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	2228
	gi 6760369	1926	GHLFAMESSSGEEYYVASDNTGTPLAVFSINGLMIKQLQYTAYGEIYYDSNPDFQMVIGF	1985
25		 2470 2480 2490 2500 2510 2520	
	NOV37	2397	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2456
30	gi 7657415	2402	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2461
	gi 9910320	2403	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2462
	gi 10241574	2440	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2499
	gi 5307761	2229	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2288
	gi 6760369	1986	HGGLYDPLTKLVHFTQRDQYDVLAGRWTSPDYTMWKNVKGEPAPFNLYMFKNNPLSSELD	2045
35		 2530 2540 2550 2560 2570 2580	
	NOV37	2457	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2516
40	gi 7657415	2462	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2521
	gi 9910320	2463	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2522
	gi 10241574	2500	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2559
	gi 5307761	2289	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2348
	gi 6760369	2046	LKNYVTDVKSWSLVMFGFQLSNIIIPGFPRAKMYFVPPPYELSESQASENGQLITGAHOTTE	2105
45		 2590 2600 2610 2620 2630 2640	
	NOV37	2517	RHNOAFMALEG-OVITKKLHASIREKAGHWFATITPIIGKGIMFAIK-EGRVTTCGVSSIA	2574
50	gi 7657415	2522	RHNOAFMALEG-OVITKKLHASIREKAGHWFATITPIIGKGIMFAIK-EGRVTTCGVSSIA	2579
	gi 9910320	2523	RHNOAFMALEG-OVITKKLHASIREKAGHWFATITPIIGKGIMFAIK-EGRVTTCGVSSIA	2580
	gi 10241574	2560	RHNOAFMALEG-OVITKKLHASIREKAGHWFATITPIIGKGIMFAIK-EGRVTTCGVSSIA	2617
	gi 5307761	2349	ROAKAFSLSGKMAEVQVSRRKAGABQSWLWFATVKSILIGKGVMLAVS-CGRVQTNVLNIA	2408
	gi 6760369	2106	ROAKAFSLSGKMAEVQVSRRKAGABQSWLWFATVKSILIGKGVMLAVS-CGRVQTNVLNIA	2164
55		 2650 2660 2670 2680 2690 2700	
	NOV37	2575	SEDSRKVASVLNNAYYLDKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2634
60	gi 7657415	2580	SEDSRKVASVLNNAYYLDKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2639
	gi 9910320	2581	SEDSRKVASVLNNAYYLDKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2640
	gi 10241574	2618	TDTSRKIASVLNSAHYLEKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2677
	gi 5307761	2409	NEDCIKVVAVLNNAYYLDKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2468
	gi 6760369	2165	NEDCIKVVAVLNNAYYLDKMHYSIEGKDTHYFVKIGSADGDLVTLGTTIGRKVLESGVNV	2224
65		 2710 2720 2730 2740 2750 2760	
	NOV37	2635	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2694
70	gi 7657415	2640	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2699
	gi 9910320	2641	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2700
	gi 10241574	2678	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2737
	gi 5307761	2469	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2525
	gi 6760369	2225	TVSQPTLLVNGRTRRFTNIEFQYSTLLLSIRYGLTPDTLDEEKARVLDAQARQALGTAWA	2281

			2770	2780	2790	2800	2810	2820	
NOV37	2695	KEQQKARDGREGSRLWTEGEKQQLLSTGRVQGYEGYYVLPVEQYPELADSSSNIOFLRON							2754
gi 7657415	2700	KEQQKARDGREGSRLWTEGEKQQLLSTGRVQGYEGYYVLPVEQYPELADSSSNIOFLRON							2759
gi 9910320	2701	KEQQKARDGREGSRLWTEGEKQQLLSTGRVQGYEGYYVLPVEQYPELADSSSNIOFLRON							2760
gi 10241574	2738	KEQQKARDGREGSRVWTEGEKQQLLSTGRVQGYEGYYVLPVEQYPELADSSSNIOFLRON							2797
gi 5307761	2526	REQQRVRDGEGVRLWTEGEKROLLSSGKVLGYDGYVLSVEQYPELADSSANNIOFLROS							2585
gi 6760369	2282	REQQRVRDGEGCARLWTEGEKROLLSACKVQGYDGYVLSVEQYPELADSSANNIOFLROS							2341
NOV37	2755	EMGKR	2759						
gi 7657415	2760	EMGKR	2764						
gi 9910320	2761	EMGKR	2765						
gi 10241574	2798	EMGKR	2802						
gi 5307761	2586	EMGKR	2590						
gi 6760369	2342	EMGKR	2346						

Table 37E lists the domain description from DOMAIN analysis results against NOV37. This indicates that the NOV37 sequence has properties similar to those of other proteins known to contain this domain.

Table 37E. Domain Analysis of NOV37	
gnl Pfam pfam02068, Metallothio_PEC, Plant PEC family metallothionein.	
(SEQ ID NO:830)	
CD-Length = 77 residues, 97.4% aligned	
Score = 41.6 bits (96), Expect = 6e-04	

NOV37:	738	CSVDCGTHGVCIGG-ACRCEEWGTAACDQRVCHPRCIEHGTCKDGKCECREGWNGEHCT	796
Sbjct:	2	CDDKCGCPSPCPGGNSCRCTSGGEAGAGDQ-----EHTTC---PC-----GEHC-	42
NOV37:	797	IDGC-PDLCNGNGRCTLGQNSWQCVCQTGWRGPGCNVAMETSCA	839
Sbjct:	43	--GCNPCTCPKTQTPTGRKGRANCSC-----GAGCTCA---SCA	76

Neurestin shows homology to a neuregulin gene product, human gamma-heregulin, a Drosophila receptor-type pair-rule gene product, Odd Oz (Odz) / Ten(m), and Ten(a), suggesting a possible function in synapse formation and morphogenesis. A mouse neurestin homolog has independently been cloned as DOC4 from the NIH-3T3 cell line. Northern blot analysis showed that neurestin is highly expressed in the brain and also in other tissues at much lower levels. In situ hybridization studies showed that neurestin is expressed in many types of neurons, including pyramidal cells in the cerebral cortex and tufted cells in the olfactory bulb during development. In adults, neurestin is mainly expressed in olfactory and hippocampal granule cells, which are known to be generated throughout adulthood. Nonetheless, in adults the expression of neurestin was experimentally induced in external tufted cells during regeneration of olfactory sensory neurons. These results suggest a role for neurestin in neuronal development and regeneration in the central nervous system

Neurestin is a putative transmembrane protein whose expression is developmentally regulated in neurons. Neurestin expression pattern were examined in mitral/tufted cells in the developing rat olfactory bulb. In the main olfactory bulb, neurestin expression was segregated in the dorso-rostral area and in the ventro-caudal area, but not in between. In the accessory
5 olfactory bulb, neurestin expression was found only in the far caudal area. This area did not completely correspond to a caudal half of the vomeronasal nerve and glomerular layers positive for a G-protein Go alpha. These spatio-temporal expression patterns suggest that neurestin functions as a target recognition molecule that specifies zonal projection patterns of olfactory and vomeronasal sensory neurons

10 The NOV37 nucleic acid of the invention encoding a Ten-M2-like protein includes the nucleic acid whose sequence is provided in Table 37A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 37A while still encoding a protein that maintains its Ten-M2-like activities and physiological functions, or a fragment of such a nucleic acid. The
15 invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar
20 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the residues may be so changed.

25 The NOV37 protein of the invention includes the Ten-M2-like protein whose sequence is provided in Table 37B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 37B while still encoding a protein that maintains its Ten-M2-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 35% of the bases
30 may be so changed.

The NOV37 nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications implicated in various diseases and disorders described below and/or other pathologies. For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: brain disorders including epilepsy,

eating disorders, schizophrenia, ADD, cancer, heart disease, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders, psoriasis, colon cancer, leukemia, AIDS, thalamus disorders, metabolic disorders including diabetes and obesity, lung diseases such as asthma, emphysema, cystic fibrosis, and cancer, pancreatic disorders including pancreatic insufficiency and cancer, and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like.

NOV37 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. For example the disclosed NOV37 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV38

A disclosed NOV38 nucleic acid of 1090 nucleotides (also referred to as CG56737-01) encoding a novel Activin Beta C Chain-like protein is shown in Table 38A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 3-5 and ending with a TAG codon at nucleotides 1068-1070. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 38A, and the start and stop codons are in bold letters.

Table 38A. NOV38 Nucleotide Sequence (SEQ ID NO:145)

<p> CAATGACCTCCTCATTGCTTCTGGCCTTTCTCCTCCTGGCTCCAACCACAGTGGCCACTCCCAGAGCTGG CGGTCAAGAGAGCATCTTGGACAAGCTGCACCTCAGCGGCCAACACTGAACCGCCCTGTGTCCAGAG CTGCTTTGAGGACTGCACTGCAGCACCTCCACGGGGTCCCACAGGGGGCACTTCTAGAGGACAACAGGGA ACAGGAATGTGAAATCATCAGCTTTGCTGAGACAGACTCCACTTCAGCCTACAGTCCCTGTCTCACTTTT CACCTGTCCACTCCTCGGTCCACCACCTGTACCATGCCCCGCTGTGGCTGCACGTGCTCCCCACCTTC CTGGCACTCTTTGCTTGAAGATCTTCGATGGGGACCAAGGAGGAGGCGCCAAGGTCCCGCACTCTCCT GGCTGAGCACCACATCACCAACCTGGGCTGGCATACTTAACCTCTGCCCTCTAGTGGCTTGAGGGGTGAG AAGTCGGGTGCTGAACTGCAACTAGACTGACAGACCCCTAGAAGGCAACAGCACAGTTACTGGACAAC CGAGGCGGCTCTTGGACACAGCAGGACACCAGCAGCCCTTCTAGAGCTTAAGATCCGAGCCAATGAGCC TGGAGCAGGCCGGGCCAGGAGGAGGACCCCACTGTGAGCTGCGACCCCTTATGTTGCGAGGCGAGAC CATTACGTAGACTTCCAGGAAGTGGGATGGCGGACTGGATACTGCAGCCGAGGGGTACCAGCTGAATT ACTCAGTGGGCACTGCCCTCCCCACCTGGCTGGCAGCCAGGCATTGCTGCCTTTTCCATTCTGCCGT CTTCAGCCTCCTCAAAGCCAACAATCCTTGGCCTGCCAGTACCTCCTGTGTGTCCCTACTGCCGAAGG CCCCCTCTCTCCTCTACCTGGATCATAGTGGCAATGTGGTCAAGACGATGTGCCAGATATGGTGGTGG AGGCTGTGGCTGCAGCTTAGCAAGAGGACCTGGGGCTTTG </p>

The disclosed NOV38 nucleic acid sequence, located on chromosome 12q13.1, has 748 of 935 bases (80%) identical to a gb:GENBANK-ID:MMU96386|acc:U96386.1 mRNA from *Mus musculus* (activin beta E subunit mRNA, complete cds) ($E = 5.2e^{-120}$).

A disclosed NOV38 polypeptide (SEQ ID NO:146) encoded by SEQ ID NO:145 is 355 amino acid residues and is presented using the one-letter amino acid code in Table 38B. Signal P, Psort and/or Hydropathy results predict that NOV38 contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.5135. The most likely cleavage site for a NOV38 peptide is between amino acids 18 and 19: TVA-TP .

Table 38B. Encoded NOV38 protein sequence (SEQ ID NO:146).

```
MTSSLLLAFLLLAPTIVATPRAGGQCPACGGPTLELESQRELLLDLAKRSILDKLHLTQRPTLNRPVSRAALRTA
LQHLHGVPQCALLEDNREQECEIISFAETDSTSAYSLLTFHLSTPRSHHLYHARLWLHVLPTLPGLTCLRIFRW
GPRRRRQGSRTLLAEHHITNLGWHTLTLPSSGLRGEKSGVLKLQDCRPLEGNSTVTGQPRRLDLAGHQPFLE
LKIRANEPGAGRARRRTPTEPATPLCCRRDHYVDFQELGWRDWILQPEGYQLNYCSGQCPPHLAGSPGIAASFH
SAVFSLLKANNPWPASTSCCVPTARRPLSLLYLDHSGNVVKTDVPMVVEACGCS
```

The disclosed NOV38 amino acid sequence has 217 of 355 amino acid residues (61%) identical to, and 253 of 355 amino acid residues (71%) similar to, the 352 amino acid residue ptnr:SWISSPROT-ACC:P55103 protein from *Homo sapiens* (Human) (inhibin beta C chain precursor (activin beta-C chain)) ($E = 7.6e^{-103}$).

NOV38 is predicted to be expressed in at least ovary and liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

Possible small nucleotide polymorphisms (SNPs) found for NOV38 are listed in Table 38C.

Table 38C: SNPs			
Consensus Position	Depth	Base Change	PAF
95	14	T > C	N/A

NOV38 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 38D.

Table 38D. BLAST results for NOV38					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 13899338 ref NP_113667.1 (NM_031479)	hypothetical protein MGC4638 [Homo sapiens]	350	291/348 (83%)	307/348 (87%)	e-153

gi 14714539 gb AAH10404.1 AAH10404 (BC010404)	inhibin beta E [Mus musculus]	350	253/352 (71%)	280/352 (78%)	e-129
gi 6680453 ref NP_32408.1 (NM_008382)	inhibin beta E [Mus musculus]	350	253/352 (71%)	280/352 (78%)	e-129
gi 13929160 ref NP_114003.1 (NM_031815)	activin beta E [Rattus norvegicus]	350	250/352 (71%)	279/352 (79%)	e-125
gi 4809189 gb AAD30133.1 AF140032_1 (AF140032)	activin beta E [Rattus norvegicus]	350	248/352 (70%)	279/352 (78%)	e-124

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 38E.

Table 38E Information for the ClustalW proteins

- 1) NOV38 (SEQ ID NO:146)
- 2) gi|13899338|ref|NP_113667.1| (NM_031479) hypothetical protein MGC4638 [Homo sapiens] (SEQ ID NO:492)
- 3) gi|14714539|gb|AAH10404.1|AAH10404 (BC010404) inhibin beta E [Mus musculus] (SEQ ID NO:493)
- 4) gi|6680453|ref|NP_032408.1| (NM_008382) inhibin beta E [Mus musculus] (SEQ ID NO:494)
- 5) gi|13929160|ref|NP_114003.1| (NM_031815) activin beta E [Rattus norvegicus] (SEQ ID NO:495)
- 6) gi|4809189|gb|AAD30133.1|AF140032_1 (AF140032) activin beta E [Rattus norvegicus] (SEQ ID NO:496)

		10	20	30	40	50	60	
NOV38	1	58
gi 13899338	1	MR	PDVQLWLVLLWALVRACGTGS	--	VCPS	CGGSKLAPQ	ALVLELAKQ	58
gi 14714539	1	MK	LPAQLWLVLLWALVWVQSTRS	--	ACPS	CGGPTLAPQ	GERALVLELAKQ	58
gi 6680453	1	MK	LPAQLWLVLLWALVWVQSTRS	--	ACPS	CGGPTLAPQ	GERALVLELAKQ	58
gi 13929160	1	MGL	SNVQLWLVLLWALVWVQSTRS	--	ACPS	CGGPTLAPQ	GERALVLELAKQ	58
gi 4809189	1	MGL	SNVQLWLVLLWALVWVQSTRS	--	ACPS	CGGPTLAPQ	GERALVLELAKQ	58
		70	80	90	100	110	120	
NOV38	59	QR	PTLNRPVSRAALRTALQHLHGVP	Q	GALL	EDNRE	QECETISFA	116
gi 13899338	59	SR	PRITHPPQAAALTRALRRLO	--	PG	SVAPGN	CEE--VIS	111
gi 14714539	59	SR	PRITRPLPQAAALTRALRRLO	--	PK	SMVPGN	REK--VIS	112
gi 6680453	59	SR	PRITRPLPQAAALTRALRRLO	--	PK	SMVPGN	REK--VIS	112
gi 13929160	59	SR	PRITRPLPQAAALTRALRRLO	--	PR	SMVPGN	REK--VIS	112
gi 4809189	59	SR	PRITRPLPQAAALTRALRRLO	--	PR	SMVPGN	REK--VIS	112
		130	140	150	160	170	180	
NOV38	117	HL	STPRSHHLYHARLWLHVLP	TL	PGTL	CLRIF	RWGPRRR	176
gi 13899338	112	HL	STPRSHHLYHARLWLHVLP	TL	PGTL	CLRIF	RWGPRRR	171
gi 14714539	113	QL	SPLWSHHLYHARLWLHVPP	S	FP	PTLYLR	IFRCGTT	171
gi 6680453	113	QL	SPLWSHHLYHARLWLHVPP	S	FP	PTLYLR	IFRCGTT	171
gi 13929160	113	QL	SPLWSHHLYHARLWLHVPP	S	FP	PTLYLR	IFRCGTT	171
gi 4809189	113	QL	SPLWSHHLYHARLWLHVPP	S	FP	PTLYLR	IFRCGTT	171
		190	200	210	220	230	240	
NOV38	177	TL	PSSGLRGEKSGVVKLQ	LD	CRPL	EGNST	VTGQPR	236
gi 13899338	172	TL	PSSGLRGEKSGVVKLQ	LD	CRPL	EGNST	VTGQPR	231
gi 14714539	172	TL	PSSGLRSEDSGVVVKLQ	LE	FRPL	DNSTA	AGLPR	231
gi 6680453	172	TL	PSSGLRSEDSGVVVKLQ	LE	FRPL	DNSTA	AGLPR	231
gi 13929160	172	TL	PSSGLRSEDSGVVVKLQ	LE	FRPL	DNSTA	AGLPR	231
gi 4809189	172	TL	PSSGLRSEDSGVVVKLQ	LE	FRPL	DNSTA	AGLPR	231

Activins are homo- or heterodimers of related beta subunits (see 147290) while inhibins are dimers composed of an alpha subunit (147380) and an activin beta subunit (summarized in Schmitt et al., Genomics 1996, 32(3):358-66). Activin proteins belong to the TGF-beta superfamily (see 190180), the members of which have important roles in cell determination, differentiation, and growth.

TGFB is a multifunctional peptide that controls proliferation, differentiation, and other functions in many cell types. It was first identified by its ability to cause phenotypic transformation of rat fibroblasts. TGFB is chemically distinct from TGFA. It has essentially no sequence homology with TGFA or with epidermal growth factor, of which TGFA is an analog. Members of the same gene family as TGFB include inhibin, which inhibits pituitary secretion of follicle stimulating hormone, and Mullerian inhibitory substance, which is produced by the testis and is responsible for regression of the Mullerian ducts (anlagen of the female reproductive system) in the male embryo. Many cells synthesize TGFB and almost all of them have specific receptors for this peptide. Alpha and beta TGFs are classes of transforming growth factors. TGFB acts synergistically with TGFA in inducing transformation. It also acts as a negative autocrine growth factor.

TGF-beta plays an important role in wound healing. A number of pathologic conditions, such as idiopathic pulmonary fibrosis, scleroderma, and keloids, which share the characteristic of fibrosis, are associated with increased TGF-beta-1 expression.

The disclosed NOV38 nucleic acid of the invention encoding a Activin Beta C Chain-like protein includes the nucleic acid whose sequence is provided in Table 38A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 38A while still encoding a protein that maintains its Activin Beta C Chain-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic

Table 39B. Encoded NOV39a protein sequence (SEQ ID NO:148).

MRLPDVQLWLVLWALVRAQGTGSVCPSGGSKLAPQAERALVLELAKQQILDGLHLTSRPRITHPPPPQAAALTRA
LRRQLQPGSVAPGNGEVVSFATVTDSTSAISSLLTFHLSTPRSHHLYHARLWLHVLPTLPGLTCLRIFRWGPRRR
RQGSRTLLAEHHITNLGWHITLTPSSGLRGEKSGVLKLQIDCRPLEGNSTVTGQPRRLDLAGHQPFLELKIRA
NEPGAGRARRRTPTEPATPLCCRRDHYVDFQELGWRDWILQPEGYQLNYCSGQCPPHPAGSPGIAASFHSAVFS
LLKANNPWPASTSCCVPTARRPLSLLYLDHNGNVKTDVPMVVEACGCS

The disclosed NOV39a amino acid sequence has 287 of 350 amino acid residues (82%) identical to, and 301 of 350 amino acid residues (86%) similar to, the 350 amino acid residue ptnr:SWISSPROT-ACC:O08717 protein from *Mus musculus* (Mouse) (Inhibin Beta E Chain Precursor (Activin Beta-E Chain)) ($E = 5.0e^{-154}$).

NOV39a is predicted to be expressed in at least the following tissues: ovary and liver. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the NOV39a sequence.

NOV39b

A disclosed NOV39b nucleic acid of 1112 nucleotides (also referred to as CG56647-03) encoding a novel Inhibin Beta E Chain-like protein is shown in Table 39C. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 40-42 and ending with a TAG codon at nucleotides 1090-1092. Putative untranslated regions, if any, found upstream from the initiation codon and downstream from the termination codon are underlined in Table 39C, and the start and stop codons are in bold letters.

Table 39C. NOV39b Nucleotide Sequence (SEQ ID NO:149)

CCATCCGAGGCTCCTGAACCGGGGCCATTACCAGGAGC**ATG**CGGCTCCC**TGATGTCCAGCTCTGGCTGG**
TGCTGCTGTGGGCACTGGTGCGAGCACAGGGGACAGGGTCTGTGTGTCCCTCCTGTGGGGGCTCCAAACT
GGCACCCCAAGCAGAACGAGCTCTGGTGCTGGAGCTAGCCAAGCAGCAAACTCCTGGATGGGTTGCACCTG
ACCACTCGTCCCAGAATAACTCATCTTCCACCCAGGCAGCGCTGACCAGAGCCCTCCGGAGACTACAGC
CAGGGAGTGTGGCTCCAGGGAATGGGGAGGAGGTATCAGCTTTGCTACTGTACAGACTCCACTTCAGC
CTACAGCTCCCTGCTCACTTTTCACTGTCCACTCCTCGGTCCCACCACCTGTACCATGCCCGCCTGTGG
CTGCACGTGCTCCCCACCCTTCTTGGCACTCTTTGCTTGAGGATCTTCCGATGGGGACCAAGGAGGAGGC
GCCAAGGGTCCCGCACTCTCCTGGCTGAGCACACATCACCAACCTGGGTGGCATACTTAACTCTGCC
CTCTAGTGGCTTGAGGGGTGAGAAGTCTGGTGTCTGAAACTGCAACTAGACTGCAGACCCCTAGAAGGC
AACAGCACAGTTACTGGACAACCGAGGCGGCTCTTGGACACAGCAGGACACCAGCAGCCCTTCCTAGAGC
TTAAGATCCGAGCCAATGAGCCTGGAGCAGGCCGGGCCAGGAGGGGACCCCACTGTGAGCCCGCGAC
CCCCTTATGTTGCGAGCGAGACCATTACGTAGACTTCCAGGAAGTGGGATGGCGGGACTGGATACCTGCAG
CCCAGGGGTACCAGCTGAATTACTGCAGTGGGCAGTGCCCTCCCCACCTGGCTGGCAGCCAGGCATTG
CTGTCTCTTCCATTCTGCCGTCTTCAGCCTCCTCAAAGCCAACAATCCTTGGCCTGCCAGTACCTCCTG
TTGTGTCCTACTGCCCGAAGGCCCTCTCTCTCTCTACCTGGATCATAATGGCAATGTGGTCAAGACG
GATGTGCCAGATATGGTGGTGGAGGCCTGTGGCTGCAGC**TAG**CAAGAGGACCTGGGGCTTTG

The disclosed NOV39b nucleic acid sequence, located on chromosome 7p13-15, has 920 of 1110 bases (82%) identical to a gb:GENBANK-ID:MMU96386|acc:U96386.1 mRNA from *Mus musculus* (activin beta E subunit mRNA, complete cds) ($E = 3.7e^{-164}$).

A disclosed NOV39b polypeptide (SEQ ID NO:150) encoded by SEQ ID NO:149 is 350 amino acid residues and is presented using the one-letter amino acid code in Table 39D. Signal P, Psort and/or Hydropathy results predict that NOV39b contains a signal peptide and is likely to be localized extracellularly with a certainty of 0.3700. The most likely cleavage site for a NOV39b peptide is between amino acids 19 and 20: VRA-QG.

Table 39D. Encoded NOV39b protein sequence (SEQ ID NO:150).					
MRLPDVQLWLVLWLVRAQGTG	SVCPSCGGSKLAPQAER	ALVLELAKQQILDGLHL	TSRPRITHLPPQAALTRA		
LRRLQPGSVAPGN	GEVVISFATVTDST	SAYSSLLTFHLSTPR	SHHLYHARLWLHVL	PTLP	GTLC
RQGSRTLLAEHHIT	NLGWHTLTLPSSGLR	GEKSGVLKLQDCR	PLEGNSTVTGQPR	RLD	TAGHQQPFLEKIRA
NEPGACRARRGTPT	CEPATPLCCRRDHY	VDFQELGWRDWILQ	PEGYQLNYCSGQC	PPHLAGSPG	IAVSFHS
LLKANNFPW	PASTSCCVPTARRPL	SLLYLDHNGNVVKT	DPDMVVEACGCS		

The disclosed NOV39b amino acid sequence has 285 of 350 amino acid residues (81%) identical to, and 299 of 350 amino acid residues (85%) similar to, the 350 amino acid residue ptnr:SWISSPROT-ACC:O08717 protein from *Mus musculus* (Mouse) (Inhibin Beta E Chain Precursor (Activin Beta-E Chain)) ($E = 1.5e^{-152}$).

NOV39b is predicted to be expressed in at least the following tissues: ovary and liver. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the NOV39b sequence.

Possible small nucleotide polymorphisms (SNPs) found for NOV39a are listed in Table 39E.

Table 39E: SNPs			
Consensus Position	Depth	Base Change	PAF
1095	11	T > C	N/A

Possible small nucleotide polymorphisms (SNPs) found for NOV39b are listed in Table 39F.

Table 39F: SNPs			
Consensus Position	Depth	Base Change	PAF
933	9	C > T	0.222

NOV39a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 39G.

Table 39G. BLAST results for NOV39a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect

gi 13899338 ref NP_113667.1 (NM_031479)	hypothetical protein MGC4638 [Homo sapiens]	350	349/350 (99%)	349/350 (99%)	0.0
gi 14714539 gb AAH10404.1 AAH10404 (BC010404)	inhibin beta E [Mus musculus]	350	288/351 (82%)	302/351 (85%)	e-153
gi 6680453 ref NP_032408.1 (NM_008382)	inhibin beta E [Mus musculus]	350	287/351 (81%)	301/351 (84%)	e-152
gi 13929160 ref NP_114003.1 (NM_031815)	activin beta E [Rattus norvegicus]	350	281/351 (80%)	294/351 (83%)	e-146
gi 4809189 gb AAD30133.1 AF140032_1 (AF140032)	activin beta E [Rattus norvegicus]	350	279/351 (79%)	294/351 (83%)	e-146

The homology of these sequences is shown graphically in the ClustalW analysis shown in Table 39H.

Table 39H Information for the ClustalW proteins

- 1) NOV39a (SEQ ID NO:148)
- 2) NOV39b (SEQ ID NO:150)
- 3) gi|13899338|ref|NP_113667.1| (NM_031479) hypothetical protein MGC4638 [Homo sapiens] (SEQ ID NO:497)
- 4) gi|14714539|gb|AAH10404.1|AAH10404 (BC010404) inhibin beta E [Mus musculus] (SEQ ID NO:498)
- 5) gi|6680453|ref|NP_032408.1| (NM_008382) inhibin beta E [Mus musculus] (SEQ ID NO:499)
- 6) gi|13929160|ref|NP_114003.1| (NM_031815) activin beta E [Rattus norvegicus] (SEQ ID NO:500)
- 7) gi|4809189|gb|AAD30133.1|AF140032_1 (AF140032) activin beta E [Rattus norvegicus] (SEQ ID NO:501)

			10	20	30	40	50	60		
NOV39a	1		MRLPDVQLWLVL	LLWALVRAO	GTG	SVCPSCGGSKLAPQA	ERALVLELAKQ	QIILDGLHLTSR	60	
NOV39b	1		MRLPDVQLWLVL	LLWALVRAO	GTG	SVCPSCGGSKLAPQA	ERALVLELAKQ	QIILDGLHLTSR	60	
gi 13899338	1		MRLPDVQLWLVL	LLWALVRAO	GTG	SVCPSCGGSKLAPQA	ERALVLELAKQ	QIILDGLHLTSR	60	
gi 14714539	1		MKLPLKAQLWLI	LLWALVWVQ	STRSAC	PSCGGPTLAPQ	GERALVLELAKQ	QIILEGLHLTSR	60	
gi 6680453	1		MKLPLKAQLWLI	LLWALVWVQ	STRSAC	PSCGGPTLAPQ	GERALVLELAKQ	QIILEGLHLTSR	60	
gi 13929160	1		MGLSNVQLWTIL	LLWALAWVQ	STRSAC	PSCGAPTILT	POGERALVLELAKQ	QIILEGLHLTSR	60	
gi 4809189	1		MGLSNVQLWTIL	LLWALAWVQ	STRSAC	PSCGAPTILT	POGERALVLELAKQ	QIILEGLHLTSR	60	
			70	80	90	100	110	120		
NOV39a	61		PRITHLP	PQAALTRALRR	LQPGSVAPGN	GEVISFATVTD	-STSAYSS	LLTFHLSTPRSH	119	
NOV39b	61		PRITHLP	PQAALTRALRR	LQPGSVAPGN	GEVISFATVTD	-STSAYSS	LLTFHLSTPRSH	119	
gi 13899338	61		PRITHLP	PQAALTRALRR	LQPGSVAPGN	GEVISFATVTD	-STSAYSS	LLTFHLSTPRSH	119	
gi 14714539	61		PRITRPL	PQAALTRALRR	LQPKSMVPGN	REKVISFATIDK	STSTYRSMLTF	FQLSPLWSH	120	
gi 6680453	61		PRITRPL	PQAALTRALRR	LQPKSMVPGN	REKVISFATIDK	STSTYRSMLTF	FQLSPLWSH	120	
gi 13929160	61		PRITRPL	PQAALTRALRR	LQPKSMVPGN	REKVISFATIDK	STSTYRSMLTF	FQLSPLWSH	120	
gi 4809189	61		PRITRPL	PQAALTRALRR	LQPRSMVPGN	REKVISFATIDK	STSTYRSVLT	FQLSPLWSH	120	
			130	140	150	160	170	180		
NOV39a	120		HLYHARLWLHV	LPPLPGTL	CLRIFRWG	PRRRRQGSRTLL	LAEHHITNL	GWHTLTPSSGLR	179	
NOV39b	120		HLYHARLWLHV	LPPLPGTL	CLRIFRWG	PRRRRQGSRTLL	LAEHHITNL	GWHTLTPSSGLR	179	
gi 13899338	120		HLYHARLWLHV	LPPLPGTL	CLRIFRWG	PRRRRQGSRTLL	LAEHHITNL	GWHTLTPSSGLR	179	
gi 14714539	121		HLYHARLWLHV	PPSFP	PGTLYLRIF	RCGTT	RCR-CFRT	FLAEHQT	TSSGWHALTLPSSGLR	179
gi 6680453	121		HLYHARLWLHV	PPSFP	PGTLYLRIF	RCGTT	RCR-CFRT	FLAEHQT	TSSGWHALTLPSSGLR	179
gi 13929160	121		HLYHARLWLHV	PPSFP	ATLYLRIF	CGTT	RCR-GSRT	FLAEHQT	TSSGWHALTLPSSGLR	179
gi 4809189	121		HLYHARLWLHV	PPSFP	ATLYLRIF	CGTT	RCR-GSRT	FLADYQT	TSSGWHALTLPSSGLR	179
			190	200	210	220	230	240		

5	NOV39a	180	GEKSGVILKLQLD	CRPLEGNSTVTG	OPRRLD	TAGHQ	OPFLELKIRANEPGAGRARRRTPT	239
	NOV39b	180	GEKSGVILKLQLD	CRPLEGNSTVTG	OPRRLD	TAGHQ	OPFLELKIRANEPGAGRARRRTPT	239
	gi 13899338	180	GEKSGVILKLQLD	CRPLEGNSTVTG	OPRRLD	TAGHQ	OPFLELKIRANEPGAGRARRRTPT	239
	gi 14714539	180	SEDSGVVVKLQLE	FRPLDLNSTAAGL	PRLLD	TAGQOR	PFLELKIRANEPGAGRARRRTPT	239
	gi 6680453	180	SEDSGVVVKLQLE	FRPLDLNSTAAGL	PRLLD	TAGQOR	PFLELKIRANEPGAGRARRRTPT	239
10	gi 13929160	180	SEESGVTKLQLE	FRPLDLNSTTARL	PRLLD	TAGQOR	PFLELKIRANEPGAGRARRRTPT	239
	gi 4809189	180	SEESGVTKLQLE	FRPLDLNSTTARL	PRLLD	TAGQOR	PFLELKIRANEPGAGRARRRTPT	239
			250	260	270	280	290	300
	NOV39a	240	CEPATPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHPAGSPGIAASFHSAVF	299			
	NOV39b	240	CEPATPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAVSFHS	299			
15	gi 13899338	240	CEPATPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAASFHSAVF	299			
	gi 14714539	240	CEPATPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAASFHSAVF	299			
	gi 6680453	240	CEPATPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAASFHSAVF	299			
	gi 13929160	240	CESETPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAASFHSAVF	299			
	gi 4809189	240	CESETPLCCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHLAGSPGIAASFHSAVF	299			
20			310	320	330	340	350	
	NOV39a	300	SLLKANNPWPASTSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
	NOV39b	300	SLLKANNPWPASTSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
	gi 13899338	300	SLLKANNPWPASTSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
	gi 14714539	300	SLLKANNPWPAGSSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
25	gi 6680453	300	SLLKANNPWPAGSSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
	gi 13929160	300	SLLKANNPWPAGSSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			
	gi 4809189	300	SLLKANNPWPAGSSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350			

30 Tables 39I-J list the domain description from DOMAIN analysis results against NOV39. This indicates that the NOV39 sequence has properties similar to those of other proteins known to contain this domain.

Table 39I Domain Analysis of NOV39

gnl|Smart|smart00204, TGFB, Transforming growth factor-beta (TGF-beta) family; Family members are active as disulphide-linked homo- or heterodimers. TGFB is a multifunctional peptide that controls proliferation, differentiation, and other functions in many cell types. (SEQ ID NO:833)
CD-Length = 102 residues, 100.0% aligned
Score = 133 bits (335), Expect = 1e-32

35	NOV39:	247	CCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQC	PPHPAGSPGIAASFHSAVFSLLKANN	306
	Sbjct:	1	CRRHDLVDFKDLGWDDWIIAPKGYNAYYCEGEC	PFPLSERLN--ATNHAIVQSLVHALD	58
40	NOV39:	307	PWPASTSCCVPTARRPLSLLYLDHNGNVVKT	DVPDMVVEACGCS	350
	Sbjct:	59	PGAVPKCCVPTKLSPLSMLYDDGNVVL	RNYPNMVVEECGCR	102

Table 39J. Domain Analysis of NOV39

gnl|Pfam|pfam00019, TGF-beta, Transforming growth factor beta like domain. (SEQ ID NO:834)
CD-Length = 105 residues, 97.1% aligned
Score = 114 bits (286), Expect = 7e-27

NOV46: 247 CCRRDHYVDFQELGWRDWILOPEGYQLNYCSGQCPHPAGSPGIAASFHSAVFSLLKANN 306
 | | |||++||| |||+ ||| ||| ||| | | + + + |++ |
 Sbjct: 4 CRLRSLYVDFRDLGWGDWIIAPEGYIANYCSGSCPFPLRDDLN--LSNHAILQTLVRLRN 61

5 NOV46: 307 PWPASTSCCVPTARRPLSLLYLDHNGNVKTDVDPDMVVEACGCS 350
 | ||| |||+||| ||| ||| | + | + | |||
 Sbjct: 62 PRAVPQPCCVPTKLSPLSMYLDNNSNVLRRLYPNMSVKECGCR 105

Activins are homo- or heterodimers of related beta subunits (see 147290) while
 10 inhibins are dimers composed of an alpha subunit (147380) and an activin beta subunit
 (summarized in Schmitt et al., Genomics 1996, 32(3):358-66). Activin proteins belong to the
 TGF-beta superfamily (see 190180), the members of which have important roles in cell
 determination, differentiation, and growth.

TGFB is a multifunctional peptide that controls proliferation, differentiation, and other
 15 functions in many cell types. It was first identified by its ability to cause phenotypic
 transformation of rat fibroblasts. TGFB is chemically distinct from TGFA. It has essentially no
 sequence homology with TGFA or with epidermal growth factor, of which TGFA is an
 analog. Members of the same gene family as TGFB include inhibin, which inhibits pituitary
 secretion of follicle stimulating hormone, and Mullerian inhibitory substance, which is
 20 produced by the testis and is responsible for regression of the Mullerian ducts (anlagen of the
 female reproductive system) in the male embryo. Many cells synthesize TGFB and almost all
 of them have specific receptors for this peptide. Alpha and beta TGFs are classes of
 transforming growth factors. TGFB acts synergistically with TGFA in inducing
 transformation. It also acts as a negative autocrine growth factor.

25 TGF-beta plays an important role in wound healing. A number of pathologic
 conditions, such as idiopathic pulmonary fibrosis, scleroderma, and keloids, which share the
 characteristic of fibrosis, are associated with increased TGF-beta-1 expression.

The disclosed NOV39 nucleic acid of the invention encoding a Activin Beta C Chain-
 like protein includes the nucleic acid whose sequence is provided in Table 39A, 39C or a
 30 fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose
 bases may be changed from the corresponding base shown in Table 39A, or 39C while still
 encoding a protein that maintains its Activin Beta C Chain-like activities and physiological
 functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids
 whose sequences are complementary to those just described, including nucleic acid fragments
 35 that are complementary to any of the nucleic acids just described. The invention additionally
 includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures
 include chemical modifications. Such modifications include, by way of non-limiting example,
 modified bases, and nucleic acids whose sugar phosphate backbones are modified or

derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, in one embodiment up to about 17% of the NOV39a residues
5 may be so changed, and in an additional embodiment up to about 18% of the NOV39b residues may be so changed.

The disclosed NOV39 protein of the invention includes the Activin Beta C Chain-like protein whose sequence is provided in Table 39B, or 39D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding
10 residue shown in Table 39B, or 39D while still encoding a protein that maintains its Activin Beta C Chain-like activities and physiological functions, or a functional fragment thereof. In one embodiment a mutant or variant protein of NOV39a, up to about 39% of the bases may be so changed.

The above defined information for this invention suggests that these Activin Beta C Chain-like proteins (NOV39) is a member of a "Activin Beta C Chain family". Therefore, the NOV39 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The nucleic acids and proteins of NOV39 are useful in Alzheimer disease-5, Myxoid liposarcoma, Stickler syndrome, type I (3), SED, Alpha-ketoglutarate dehydrogenase
25 deficiency, Cerebral cavernous malformations-2, Greig cephalopolysyndactyly syndrome, Hyperinsulinism, familial, MODY, type 2, Pallister-Hall syndrome, Polydactyly, postaxial, types A1 and B, Polydactyly, postaxial, type IV, Retinitis pigmentosa-9, Charcot-Marie-Tooth neuropathy-2D, Colton blood group, Deafness, autosomal dominant 5, Macular dystrophy, dominant cystoid, Radioulnar synostosis with amegakaryocytic thrombocytopenia, Von
30 Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, fertility and/or other pathologies and disorders.

NOV39 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods

known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. For example the disclosed NOV39 protein have multiple hydrophilic regions, each of which can be used as an immunogen. This novel protein also has value in development of powerful assay system for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV40

A disclosed NOV40 nucleic acid of 1606 nucleotides (also referred to as CG56097-01) encoding a UDP glycosyltransferase-like protein is shown in Table 40A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 1-3 and ending with a TAG codon at nucleotides 1600-1602. The start and stop codons are shown in bold in Table 40A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 40A. NOV40 nucleotide sequence (SEQ ID NO:151).

ATGGCTATGAAATGGACTTCAGTCCTTCTGTTGATACAGCTGAGCTATTACTCTAGCTCTGGGAGTTGTGGA
AATGTGCCGCTGTGGCCCATGGAATATAGTCCTTGGATGAATATAAGACAATCCTGGATAAACTTATGCAG
ATAAGTCATGAGGTGACTGTTCTAACATTGTCAGCTCCATTCTTGTGATCCCAACATAACATCTGTTACT
AAATTTGAGGTTTATTCTATATCTGTAATTAAGATGATTTTGCAGGGTTTTTTTTCACACAACAGATTACT
AAATGGATACATGATCTTCCAAAACATATATTTGGTTTAAATGTGTCCCTTCAAGAATATTCTTTGGGAA
TATTTCTGGTTATACTGAGAAGTCTTTAAAGATGATGTTTGAACAAGAAACTTATGACAAACCTACAGAA
TCAAGGTCTGATGTCGTTTCATGCAAATGCCATTGGTCCCTTGGAGAGCTGCTGGCTGAGCTATTAATAATA
TCCTTTGTGTACAGTCTCCACTTCTCTCGGCTACACATTTGAGAAATACAGTGGAGGATTTCTACTTCCA
CCTTCCTATGGAGCTGTATTCTGTCTCAGAAATTAAGTGGTTCGATGACATTCATGGAGACAGTAAGAAATATT
ATATATGTGTTTATTTTGACTTTTGGTTCCAAACATTTGATATGAAGAAGGGAGACCAGTTTACAGTGAA
GTTCTAGGTAAGTCATGTTTATCTGAGATAATGGGAAAAGCTGAAATGTGGCTCATTGCAAACTACTGG
TATTTGGAATTTCTCGCCCACTCTTACCTAATTTTGAATTTGTTGTAAGACTCTACTGCAAACTGTCAAC
CCCCTGCCTAAGGAGAAAATGGAAGAATTTGCCAGAGCTCTGATGAAGACGGTGTTGTGTTTCTCTGGAG
TCAGCTGTGCAAACTTACAGAGAAGAAAAGCTGATCTTATCACTTCGGCCCTGGCTCAGATTTCCACAAAAA
GTCATGAAGTTCGGAAGGAAACAAATACCTTAAGATCCAATACCTCAGTGGCATAGGTCATCCACAGAAT
GAATGTCTTATCCTAGATCATCCCAAACCAAGCCTTTATACTTATGGTGAACAAATAGCATCTATGAG
ATGATCTACCGTGGAGTCCCTTCCATGGGCATTCTTTGTTTGGGACCAACATGATAACATTGCTCACATG
AAGGCCAAGGGAGCAGCTGTTATATTGGACTTGAGCACAAAGTCAAGTACAGATTGCTCGATATATCTGTG
TTCGTATCTTTATTTTATCTTTCAGATATAAAGAGAGTGTATGAAATTATCAAGAATTCAACATGATCAA
CCAGTGAAGCCCCTGGATCGAGCAGTCTTCTGGATTGAATTTGTATGCGCCACAAAGGAGCCAAACACCTT
CGAGTTGCAGCCCCTGACCTCACCTGGTTCCAGTACCCTCTTTGGATGTGATTGGGTTTCTGCTGGCCTGT
GTGGCAACTGTGACATTTATCATCACAAAGTGTGTCTGTTTGTCTGGAAGTTTACTAGAAAAGTGAAG
AAGGAAAAAGGGATTAGTTAT

In a search of public sequence databases, the NOV40 nucleic acid sequence, located on chromosome 4, has 1305 of 1606 bases (81%) identical to a gb:GENBANK-ID:HUMUDPGTA|acc:J05428.1 mRNA from *Homo sapiens* (Human 3,4-catechol estrogen UDP-glucuronosyltransferase mRNA, complete cds) ($E = 6.4e^{-217}$).

The disclosed NOV40 polypeptide (SEQ ID NO:152) encoded by SEQ ID NO:151 has 533 amino acid residues and is presented in Table 40B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV40 has no signal peptide and is likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.8200.

Alternatively, NOV40 may also localize to the plasma membrane with a certainty of 0.4600, to the microbody (peroxisome) with a certainty of 0.3012, or to the lysosome (membrane) with a certainty of 0.2000. The most likely cleavage site for NOV40 is between positions 20 and 21: SSS-GS.

5

Table 40B. Encoded NOV40 protein sequence (SEQ ID NO:152).

```
MAMKWTSVLLLIQLSYSSSGSCGNVPLWPMEYSPWMNIKTILDKLMQISHEVTVLTLASILVDPNITSVT
KFEVYSISVIKDDFAGFFFTQQITKWIHDLPKHIFWFKCVFPKNILWEYSGYTEKFFKDVVLNKKLMTNLQE
SRSDVVHANAIGPFGELLAELLKISFVYSLHFSPGYTFEKYSGGFLLPPSYGAVILSELSGSMTFMETVRNI
IYVFYDFWFQTFDMKKGDQFYSEVLGKSCFLSEIMGKAEMLIRNYWYLEFPRLLPNFEFVRLYCKPVN
PLPKEMEEFAQSSDEGDGVVSLESVAQNLTEEKADLITSALAQIPQKVMKFGKRKPNLRSNTQWHRWIPQN
ECLILDHPQTKAFITYGGTNSIYEMIYRGVPSMGIPLFADQHDNIAHMKAKGAIVLDLSTKSSTDLLDISV
FVSLFLSFRYKESVMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGAKHLRVAARDLTWFAQYHSLDVIGFLLAC
VATVTFIITKCCLFCFWKFRKVKKEKRD
```

A search of sequence databases reveals that the NOV40 amino acid sequence has 353 of 533 amino acid residues (66%) identical to, and 412 of 533 amino acid residues (77%) similar to, the 529 amino acid residue ptnr:SWISSPROT-ACC:P16662 protein from *Homo sapiens* (Human) (UDP-Glucuronosyltransferase 2b7 Precursor, Microsomal (EC 2.4.1.17) (UDPGT) (3,4-Catechol Estrogen Specific) (UDPGTH-2)) ($E = 7.2e^{-185}$).

NOV40 is predicted to be expressed in at least the following tissues: liver tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV40 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 40C.

Table 40C. BLAST results for NOV40

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 4507825 ref NP_01065.1 (NM_001074)	UDP glycosyltransferase 2 family, polypeptide B7	529	353/536 (65%)	412/536 (76%)	0.0
gi 6175083 sp P06133 UDB4_HUMAN	UDP- GLUCURONOSYLTRANSFERASE 2B4 PRECURSOR, MICROSOMAL (UDPGT) (HYODEOXYCHOLIC ACID) (HLUG25) (UDPGTH-1)	528	355/536 (66%)	409/536 (76%)	0.0
gi 484383 pir JN0619	glucuronosyltransferase (EC 2.4.1.17) 2B-4 precursor - human	528	354/536 (66%)	408/536 (76%)	0.0

gi 3153832 gb AAC95002.1 (AF064200)	UDP-glucuronosyltransferase 2B4 precursor [<i>Homo sapiens</i>]	528	354/536 (66%)	409/536 (76%)	0.0
gi 4079707 gb AAC98726.1 (AF016310)	UDP-glucuronosyltransferase [<i>Macaca fascicularis</i>]	529	351/536 (65%),	410/536 (76%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 40D. In the ClustalW alignment of the NOV40 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 40D. ClustalW Analysis of NOV40

- 1) Novel NOV40 (SEQ ID NO:152)
- 2) gi|4507825|ref|NP_001065.1| (NM_001074) UDP glycosyltransferase 2 family, polypeptide B7 (SEQ ID NO:502)
- 3) gi|6175083|sp|P06133|UDB4_HUMAN UDP-GLUCURONOSYLTRANSFERASE 2B4 PRECURSOR, MICROSOAL (UDPGT) (HYODEOXYCHOLIC ACID) (HLUG25) (UDPGTH-1) (SEQ ID NO:503)
- 4) gi|484383|pir|JN0619 glucuronosyltransferase (EC 2.4.1.17) 2B-4 precursor - human (SEQ ID NO:504)
- 5) gi|3153832|gb|AAC95002.1| (AF064200) UDP-glucuronosyltransferase 2B4 precursor [*Homo sapiens*] (SEQ ID NO:505)
- 6) gi|4079707|gb|AAC98726.1| (AF016310) UDP-glucuronosyltransferase [*Macaca fascicularis*] (SEQ ID NO:506)

			10	20	30	40	50	60	
NOV40	1	MSMKWTSVLLLIQLSYSSSGSCGNVPLWPMES	SPWMNIKTILD	KLM	QISHEVTVL	TL	SA	60	
gi 4507825	1	MSVWKTSVLLLIQLSFCFSSGNC	CGKVLVWAAEY	SHWMNIKTILDE	LH	ORGHEVTVL	ASSA	60	
gi 6175083	1	MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTE	FSHWMNIKTILDE	LV	QRGHEVTVL	ASSA	60		
gi 484383	1	MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTE	FSHWMNIKTILDE	LV	QRGHEVTVL	ASSA	60		
gi 3153832	1	MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTE	FSHWMNIKTILDE	LV	QRGHEVTVL	ASSA	60		
gi 4079707	1	MSVWKTSVLLLIQLSFCFSSGNC	CGKVLVWAAEY	SHWMNIKTILDE	LH	ORGHEVTVL	ASSA	60	
			70	80	90	100	110	120	
NOV40	61	SILVDPNITS	VIKFEVYS	TSVIKDD	FAGFFFT	QOTIKWIH	DLPKHIF	FWFKCV	PKN
gi 4507825	61	SILFDPNNS	SALKIEI	YPTSL	TKTELEN	-FTMCOIKRWS	-DLPKDTFW	LYFSQV	QEIMSI
gi 6175083	61	SISFDPNSP	STLKFEVYP	VS	SLTKTEFED	-IIKOLVKRWA	-ELPKDTFWS	YFSQVQ	QEIMWT
gi 484383	61	SISFDPNSP	STLKFEVYP	VS	SLTKTEFED	-IIKOLVKRWA	-ELPKDTFWS	YFSQVQ	QEIMWT
gi 3153832	61	SISFDPNSP	STLKFEVYP	VS	SLTKTEFED	-IIKOLVKRWA	-ELPKDTFWS	YFSQVQ	QEIMWT
gi 4079707	61	SILFDPNNS	SALKIEI	VEFPT	SLTKTEFEN	-IIKQOIKRWS	-ELPKDTFW	LYFSQV	QOIMWK
			130	140	150	160	170	180	
NOV40	121	YSGYTEKFEK	DVVLNKKLM	TNLOESRS	SDVVHANAT	GPF	GELLA	LAELLKIS	FFVYSIH
gi 4507825	119	FGDITRK	FKCDVVS	NKKFMKK	VOESRFDV	TFADAI	FP	CELLA	ELFNIP
gi 6175083	119	FNDILRK	FKCDIVS	NKKLMKK	LQESRFDV	VLADAV	FP	FGELLA	LAELLKIP
gi 484383	119	FNDILRK	FKCDIVS	NKKLMKK	LQESRFDV	VLADAV	FP	FGELLA	LAELLKIP
gi 3153832	119	FNDILRK	FKCDIVS	NKKLMKK	LQESRFDV	VLADAV	FP	FGELLA	LAELLKIP
gi 4079707	119	FGDITRK	FKCDVVS	NKKLMKK	LQESRFDV	VFADAI	FP	CELLA	ELLNTP

			190	200	210	220	230	240	
5	NOV40	181	TFEKYSGGFLPPSYGAVILSELSSGSMTFMETVNTIYVYEDFWFQIFDMKKGDQFYSE	240					
	gi 4507825	179	TFEKHSGGFIFFPSYVPVVMSELSDQMTFMERVKNMIIYVLYDFWFQIFDMKKWDQFYSE	238					
	gi 6175083	179	AIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIIYVLYDFWFQIFDMKKWDQFYSE	238					
	gi 484383	179	AIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIIYVLYDFWFQIFDMKKWDQFYSE	238					
	gi 3153832	179	AIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIIYVLYDFWFQIFDMKKWDQFYSE	238					
10	gi 4079707	179	NFEKHCGGFLPPSYVPVVMSELSDQMTFMERVKNMIIYVLYDFWFQIFDMKKWDQFYSE	238					
			250	260	270	280	290	300	
15	NOV40	241	VLGKSCFLSEIMCKAEMWLIRNYWYLEFPRPLLPNFEFVRLVCKPVNPLPKEMEEFVQ	300					
	gi 4507825	239	VLGRPTTLSETMGKADVWLIRNSWNFQFPHPLPNVDFVGLHCKPAKPLPKE-MEEFVQ	297					
	gi 6175083	239	VLGRPTTLSETMAKADIWLIRNYWDFQFPHPLPNVEFVGLHCKPAKPLPKE-MEEFVQ	297					
	gi 484383	239	VLGRPTTLSETMAKADIWLIRNYWDFQFPHPLPNVEFVGLHCKPAKPLPKE-MEEFVQ	297					
	gi 3153832	239	VLGRPTTLSETMAKADIWLIRNYWDFQFPHPLPNVEFVGLHCKPAKPLPKE-MEEFVQ	297					
	gi 4079707	239	VLGRPTTLSETMGKADIWLIRNSWNFQFPHPLPNVDFVGLHCKPAKPLPKE-MEEFVQ	297					
20			310	320	330	340	350	360	
	NOV40	301	SSDEEG-VVFSLESAAVONLTEEKADLITSALACIPQKVMK--FGRKPNTLRSTQWHRWI	357					
	gi 4507825	298	SSGENGVVVFSLGSMVSNMTEERANVIASALACIPQKVLWRFDGKPDITGLNTRLYKWI	357					
	gi 6175083	298	SSGENGVVVFSLGSMVSNMTEERANVIASALACIPQKVLWRFDGKPDITGLNTRLYKWI	357					
25	gi 484383	298	SSGENGVVVFSLGSMVSNMTEERANVIASALACIPQKVLWRFDGKPDITGLNTRLYKWI	357					
	gi 3153832	298	SSGENGVVVFSLGSMVSNMTEERANVIASALACIPQKVLWRFDGKPDITGLNTRLYKWI	357					
	gi 4079707	298	SSGENGVVVFSLGSMVSNMTEERANVIASALACIPQKVLWRFDGKPDITGLNTRLYKWI	357					
30			370	380	390	400	410	420	
	NOV40	358	PQNECLILDHPQTKAFITYGGTNSIYEMIYRGVPSMGIPLFADQPDNIAHMKAKGAAVSL	417					
	gi 4507825	358	PQND--LLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPDNIAHMKACAAVRV	415					
	gi 6175083	358	PQND--LLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPDNIAHMKAKGAAVSL	415					
	gi 484383	358	PQND--LLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPDNIAHMKAKGAAVSL	415					
35	gi 3153832	358	PQND--LLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPDNIAHMKAKGAAVSL	415					
	gi 4079707	358	PQND--LLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPDNIAHMKAKGAAVSL	415					
40			430	440	450	460	470	480	
	NOV40	418	DLSTKSSTDLLDISVFSVSLFLSFRYKESVMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	477					
	gi 4507825	416	DFHTMSSTDLLN--ALKRVINDESYKENVMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	473					
	gi 6175083	416	DFHTMSSTDLLN--ALKTVINDPLYKENAMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	473					
	gi 484383	416	DFHTMSSTDLLN--ALKTVINDPLYKENAMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	473					
	gi 3153832	416	DFHTMSSTDLLN--ALKTVINDPLYKENAMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	473					
45	gi 4079707	416	DFHTMSSTDLLN--ALKTVINDPLYKENAMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGA	473					
			490	500	510	520	530		
50	NOV40	478	KHLRVAARDLTWFQYHSLDVIGFLLACVATVIFITKCLFCFWKFKTRKVKKEKRD	533					
	gi 4507825	474	KHLRVAADLTWFQYHSLDVIGFLLVCVATVIFITKCLFCFWKFKARKAKKGKND	529					
	gi 6175083	474	KHLRVAADLTWFQYHSLDVIGFLLACVATVIFITKCLFCVWKFFVRTGKKGKRD	528					
	gi 484383	474	KHLRVAADLTWFQYHSLDVIGFLLACVATVIFITKCLFCVWKFFVRTGKKGKRD	528					
	gi 3153832	474	KHLRVAADLTWFQYHSLDVIGFLLACVATVIFITKCLFCVWKFFVRTGKKGKRD	528					
55	gi 4079707	474	KHLRPAADLTWFQYHSLDVIGFLLACVATVIFITKCLFCFWKFKARKGKKGKSD	529					

Table 40E lists the domain descriptions from DOMAIN analysis results against NOV40. This indicates that the NOV40 sequence has properties similar to those of other proteins known to contain this domain.

Table 40E Domain Analysis of NOV40

gnl|Pfam|pfam00201, UDPGT, UDP-glucuronosyl and UDP-glucosyl
transferase. (SEQ ID NO:835)
CD-Length = 501 residues, 100.0% aligned
Score = 587 bits (1514), Expect = 4e-169

NOV40:	24	GNVPLWPMEYSPWMNIKTILDKLMQISHEVTVLTLASILVDPNITSVTKFEVYSISVIK	83
Sbjct:	1	GKVLVWPMGDGSHWMNMKGILLELVQRGHEVTVLRPSASILIGPAKPSNLKFETYPDSATK	60
NOV40:	84	DDFAGFFFTQQITKWIHDLPKHIFWFKCVPFKNILW-----EYSGYTEKFFKDVVLNK	136
Sbjct:	61	EELLENLF-----PKRVMN-----WFMEAAEAGTVWSYFSALQEYSDGARVSCKELVGNK	109
NOV40:	137	KLMTNLQESRSDVVHANAIGPFGELLAELLKISFVYSLHFSPGYTFEKYSGGFLLPSPSYG	196
Sbjct:	110	FLMTKLQESSFDVVLADPVPWPCGALLAELLHIPTVYSLRFVPGYAAEKADGGLPAPPSYV	169
NOV40:	197	AVILSELSGSMTFMETVRNIIYVFYFDFWFQTFDMKKGDQFYSEVLGKSCFLSEIMGKAE	256
Sbjct:	170	PVRLSDLSGDMTFGERVKNMILIMLYFDFWFQRFPP-KKWDQFASSELLGRPVTLPEDLSKAS	228
NOV40:	257	MWLIRNYWYLEFFRPLLPNFEEFVVRLYCKPVNPLPEKMEFEAQSSDEDGVV-FSLESAY	315
Sbjct:	229	AWLLRNYWDLEFFRPLLPNMEFIGGLNCKPAKPLPQE-MEAFVQSSGEHGVVVFSLGSMV	287
NOV40:	316	QNLTEEKADLITSALAQIPQKVM-KF-GRKPNTLRSNTQWHRWIPQNECLILDHPQTKAF	373
Sbjct:	288	SNIPEEKANEIASALAQIPQKVLWRFDGTPSTLGNNTLRVWLPQND--LLGHPKTRAF	345
NOV40:	374	ITYGGTNSIYEMIYRGVPSMGIPLFADQHDNIAHMAKAKGAAVILDLSTKSSTDLLDISVF	433
Sbjct:	346	VTHAGSNGVYEAICHGVPMVGMPLFGDQMDNAKHMEAKGAAVTLNVLTMTSEDLLNALK-	404
NOV40:	434	VSLFLSFRYKESVMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGAKHLRVAARDLTFWQYH	493
Sbjct:	405	-TVINDPSYKENIMRLSSIHHDDQPVKPLDRAVFWIEFVMRHKGAKHLRPAAHDLTWYQYH	463
NOV40:	494	SLDVIGFLLACVATVTFTITKCLFCFWKFTRKVKKEK	531
Sbjct:	464	SLDVIGFLLACVATVAFITFKCLFGYRKFGVGGKKRKVK	501

The UDP-glucuronosyltransferases, a group of isoenzymes located primarily in hepatic endoplasmic reticulum and nuclear envelope, are encoded by a large multigene family that has evolved to produce catalysts with differing but overlapping substrate specificities. Two subfamilies are recognized by sequence identities. UGT1 consists of at least 4 isoenzymes that catalyze the glucuronidation of phenols and bilirubin. All 4 map to chromosome 2 and probably derive from the same gene (UGT1). The UGT2 family contains at least 5 members catalyzing steroid or bile acid glucuronidation. Members of the subfamily share 65 to 90% amino acid sequence identity. However, unlike the phenol UGT cDNAs, where the high degree of identity is concentrated in the 3-prime region of the cDNA, the steroid UGTs have a high degree of sequence homology throughout the cDNA. The disclosed NOV40 nucleic acid of the invention encoding a UDP Glycosyltransferase -like protein includes the nucleic acid whose sequence is provided in Table 40A or a fragment thereof. The invention also includes a

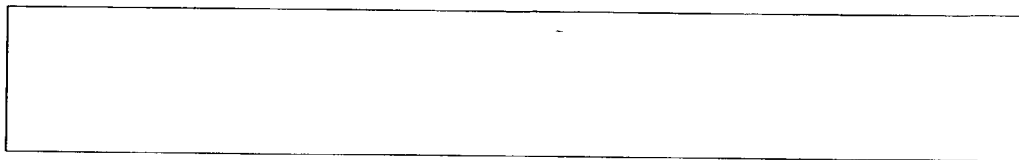
mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 40A while still encoding a protein that maintains its UDP[Glycosyltransferase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are
5 complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or
10 derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 19 percent of the bases may be so changed.

The disclosed NOV40 protein of the invention includes the UDP Glycosyltransferase -
15 like protein whose sequence is provided in Table 40B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 40B while still encoding a protein that maintains its UDP Glycosyltransferase -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 35 percent of the residues may be so changed.

20 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this UDP Glycosyltransferase -like protein (NOV40) is a member of a "UDP Glycosyltransferase family". Therefore, the NOV40 nucleic acids and proteins identified here may be useful in potential therapeutic applications
25 implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues
30 and cell types composing (but not limited to) those defined here.

The NOV40 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in brain disorders including Crigler-Najjar syndrome, Gilbert syndrome, and/or other diseases and pathologies.



In a search of public sequence databases, the NOV41a nucleic acid sequence, located on chromosome 2, has 1820 of 2156 bases (84%) identical to a gb:GENBANK-
ID:RATNHEXIV|acc:M85301.1 mRNA from *Rattus norvegicus* (Rat sodium-hydrogen
5 exchange protein-isoform 4 (NHE-4) mRNA, complete cds) ($E = 6.4e^{-217}$).

The disclosed NOV41a polypeptide (SEQ ID NO:154) encoded by SEQ ID NO:153 has 708 amino acid residues and is presented in Table 41B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV41a has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.8200. Alternatively,
10 NOV41a may also localize to the Golgi body with a certainty of 0.4600, to the endoplasmic reticulum (membrane) with a certainty of 0.3700, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The most likely cleavage site for NOV41a is between positions 26 and 27: SEA-SS.

Table 41B. Encoded NOV41a protein sequence (SEQ ID NO:154).

MALQMFVTYSPWNCLELLLVALECSEASSDLNESANSTAQYASNAWFAAASSEPEEGISVFELDYDVVQIPYE VTLWILLASLAKIGFHLVHRLPGLMPESCLLILVGALVGGIIFGTDHKSPVMDSSIYFLYLLPPIVLEGGY FMPTRPFFENIGSILWWAVLGALINALGIGLSLYLICQVKAFLGLGDNLLQNLFLGSLISAVDPVAVLAVFE EARVNEQLYMMIFGEALLNDGITVVVLYNMLIAFTKMHKFEDIETVDILAGCARFIVGLGGVLFVGFIF SAFITRFTQNISAIIEPLIVFMFSYLSYLAETLYLSGILAITACAVTMKKYVEENVQSSTYTTIKYFMKMLS SVSETLIFIFMGVSTVGKNHEWNWAFICFTLAFQCQIWRASVFAFYISNQFRTFPFSIKDQCIIFYSGVRG AGSFSLAFLPLSLFPRKKMFVTATLVVIYFTVFIQGITVGPLVRYLDVKKTNKKESINEELHIRLMDHLKA GIEDVCGHWSHYQVRDKFKKFDHRYLRKILIRKNLPKSSIVSLYKKLEMKQAIEMVETGILSSTAFSIPHQA QRIQGIKRLSPEDVESIRDILTSNMYQVRQRTLSYNKYNLKPQTSEKQAKEILIRRQNTLRESMRKGHSLPW GKPAGTKNIRYLSYPYGNPQSAGRDRTRAGFSGKLPTWLLCCFSVESGGKYLGVWAKRQH

15 A search of sequence databases reveals that the NOV41a amino acid sequence has 599 of 688 amino acid residues (87%) identical to, and 631 of 688 amino acid residues (91%) similar to, the 717 amino acid residue ptnr:SWISSPROT-ACC:P26434 protein from *Rattus norvegicus* (Rat) (Sodium/Hydrogen Exchanger 4 (NA(+)/H(+)) Exchanger 4) (NHE-4)) ($E =$
20 0.0).

NOV41a is predicted to be expressed in at least the stomach. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in stomach, colon and small intestine; lesser amounts in kidney, brain, uterus and skeletal muscle because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:RATNHEXIV|acc:M85301.1) a closely related Rat sodium-hydrogen exchange protein-isoform 4 (NHE-4) mRNA, complete cds homolog.

5 **NOV41b**

In the present invention, the target sequence identified previously, NOV41a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was
10 examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely
15 related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary
20 gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with
25 another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV41b. This differs from the previously identified sequence (NOV41a) in having 17 different aminoacids.

A disclosed NOV41b nucleic acid of 2436 nucleotides (also referred to as CG56680-
30 02) encoding a Sodium/Hydrogen Exchanger 4-like protein is shown in Table 41C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 86-88 and ending with a TAA codon at nucleotides 2369-2371. The start and stop codons are shown in bold in Table 41C, and the 5' and 3' untranslated regions, if any, are underlined.

VTLWILLASLAKIGFHLRHRLPGLMPESCILLILVGALVGGIIFGTDHKSPPVMDSSIIYFLYLLPPIVLEGGY
 FMPTRPFFENIGSILWWAVLGCALINALGIGLSLYLICQVKAFGLGDVNLLQNLFGSLISAVDPVAVLAVFE
 EARVNEQLYMMIFGEALLNDGITVVLNMLIAFTKMHKFEDIETVDILAGCARFIVVGLGGVLFGLVFGFIS
 AFITRFTQNISAIIEPLIVFMFSYLSYLAETLYLSGILAITACAVTMKKYVEENVQSQTSTTIKIFMKMLSS
 VSETLIFIFMGVSTVGKNHEWNWAFICFTLAFQCIWRAISVFALFYISNQFRTFFFSIKDQCIIFYSGVRGA
 GSFSLAFLPLSLFPRKKMFVTATLVVIYFTVFIQGITVGPLVRYLDVKKTNKKESINEELHIRLMDHLKAG
 IEDVCGHWSHYQVRDKFKKFDHRYLRKILIRKNLPKSSIVSLYKKLEMKQAIEMVETGILSSSTAFSIPHQAQ
 RIQGIKRLSPEDVESIRDILTSNMYQVRQRTLSYNKYNLKPQTSEKQAKEILIRRONTLRESMRKGHSLPWG
 KPAGTKNIRYLSYPYGNPQSAGRDRTRAAGFSGKLPWLLWLRFGRGGQLTMDTAGTITGPVILCSKKNSVI
 VHKIVLVFLKSLSSYNCERRISGIQKRAIEFAVLKLLNMDL

A search of sequence databases reveals that the NOV41b amino acid sequence has 606 of 717 amino acid residues (84%) identical to, and 641 of 717 amino acid residues (89%) similar to, the 717 amino acid residue ptnr:SWISSPROT-ACC:P26434 protein from *Rattus norvegicus* (Rat) (Sodium/Hydrogen Exchanger 4 (NA(+)/H(+) Exchanger 4) (NHE-4)) (E = 0.0).

NOV41b is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. .

NOV41a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 41E.

Table 41E. BLAST results for NOV41a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 127814 sp P26434 NAH4_RAT	SODIUM/HYDROGEN EXCHANGER 4 (NA(+)/H(+)) EXCHANGER 4) (NHE-4)	717	599/688 (87%)	631/688 (91%)	0.0
gi 1346658 sp P48763 NAH2_RAT	SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+)) EXCHANGER 2) (NHE-2) (H7)	813	421/659 (63%)	523/659 (78%)	0.0
gi 1709222 sp P50482 NAH2_RABIT	SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+)) EXCHANGER 2) (NHE-2)	809	419/659 (63%)	522/659 (78%)	0.0
gi 15529998 ref NP_003039.2 (NM_003048)	solute carrier family 9 (sodium/hydrogen exchanger), isoform 2 [<i>Homo sapiens</i>]	812	405/611 (66%)	499/611 (81%)	0.0

gi 6981560 ref NP_036785.1 (NM_012653)	solute carrier family 9 (sodium/hydrogen exchanger 2), antiporter 2, Na+/H+ (Na+/H+ exchanger 2) [Rattus norvegicus]	697	372/560 (66%)	457/560 (81%)	0.0
--	--	-----	---------------	---------------	-----

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 41F. In the ClustalW alignment of the NOV41 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 41F. ClustalW Analysis of NOV41

- 1) Novel NOV41a (SEQ ID NO:154)
- 2) Novel NOV41b (SEQ ID NO:156)
- 3) gi|127814|sp|P26434|NAH4_RAT SODIUM/HYDROGEN EXCHANGER 4 (NA(+)/H(+) EXCHANGER (NHE-4) (SEQ ID NO:507)
- 4) gi|1346658|sp|P48763|NAH2_RAT SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+) EXCHANGER (NHE-2) (H7) (SEQ ID NO:508)
- 5) gi|1709222|sp|P50482|NAH2_RABIT SODIUM/HYDROGEN EXCHANGER 2 (NA(+)/H(+) EXCHANGER 2) (NHE-2) (SEQ ID NO:509)
- 6) gi|15529998|ref|NP_003039.2| (NM_003048) solute carrier family 9 (sodium/hydrogen exchanger), isoform 2 [*Homo sapiens*] (SEQ ID NO:510)
- 7) gi|6981560|ref|NP_036785.1| (NM_012653) solute carrier family 9 (sodium/hydrogen exchanger 2), antiporter 2, Na+/H+ (Na+/H+ exchanger 2) [*Rattus norvegicus*] (SEQ ID NO:511)

		10	20	30	40	50	60		
NOV41a	1	---	MALQMFVTYSPWNC	LLLLVALECSEASSDLNES	-----	ANSTAQYASNAWFAAA	49		
NOV41b	1	---	MALQMFVTYSPWNC	LLLLVALECSEASSDLNES	-----	ANSTAQYASNAWFAAA	49		
gi 127814	1	---	MGPAMLR	AFSSWKWLLLLMVL	TCLSEASSYVNES	-----	SSPTGQQT	PDARFAAS	49
gi 1346658	1	---	MGPSTAH	RMRAPLSWLLLLLS	LOVAVPAGALAE	TLLDAPGARGASSNP	PSPASVVAPG	60	
gi 1709222	1	---	MESAGTGR	SLRTPPPRLLLLL	-LOVAGPAGALAE	TLNAPKAMGTSSSP	PLSPASVVAPG	59	
gi 15529998	1	---	MEPLGNWR	SLRAPLPMLLLLL	-LOVAGPVGALAE	TLNAPRAMGTSSSP	PSPASVVAPG	59	
gi 6981560	1	---	---	---	---	---	---	1	
		70	80	90	100	110	120		
NOV41a	50	SSEPEE	-GISVFEL	DYDYVQIPYEVT	LWILLASLAKIGF	HLHYHRLPGLMPESCLLI	IVGA	108	
NOV41b	50	SSEPEE	-GISVFEL	DYDYVQIPYEVT	LWILLASLAKIGF	HLHYHRLPGLMPESCLLI	IVGA	108	
gi 127814	50	SSDPDE	-RISVFEL	DYDYVQIPYEVT	LWILLASLAKIGF	HLHYHRLPGLMPESCLLI	IVGA	108	
gi 1346658	61	TTTFEES	RLPVFTLDYPH	VQIPFEITLWILLASLAKIGF	HLHYHKLPTITVPESCLLI	MVGL	120		
gi 1709222	60	TTTFEES	RLPVFTLDYPH	VQIPFEITLWILLASLAKIGF	HLHYHKLPTITVPESCLLI	MVGL	119		
gi 15529998	60	TTTFEES	RLPVFTLDYPH	VQIPFEITLWILLASLAKIGF	HLHYHKLPTITVPESCLLI	MVGL	119		
gi 6981560	1	---	---	---	---	---	MVGL	4	
		130	140	150	160	170	180		
NOV41a	109	LVGGIIF	GTGDKSPVMD	SSISYFLYLLPPIV	LEGGYFMPTRPFFENIGS	ILWNAVLGAL	I	168	
NOV41b	109	LVGGIIF	GTGDKSPVMD	SSISYFLYLLPPIV	LEGGYFMPTRPFFENIGS	ILWNAVLGAL	I	168	

5	gi 127814	109	LVGSIIFGTHHKSPVMDSSIFLYLLPPIVLES	168
	gi 1346658	121	LLGGIIFGVDEKSPAMKTDVFFLYLLPPIVLDAGYFMPTRPFFENIGTIFWYAVVGTILW	180
	gi 1709222	120	LLGGIIFGVDEKSPAMKTDVFFLYLLPPIVLDAGYFMPTRPFFENIGTIFWYAVVGTILW	179
	gi 15529998	120	LLGGIIFGVDEKSPAMKTDVFFLYLLPPIVLDAGYFMPTRPFFENIGTIFWYAVVGTILW	179
	gi 6981560	5	LLGGIIFGVDEKSPAMKTDVFFLYLLPPIVLDAGYFMPTRPFFENIGTIFWYAVVGTILW	64
<div>190200210220230240</div>				
10	NOV41a	169	NALGIGLSLYLICQVKAFLGCDVNLQNLFFGSLISAVDPVAVLAVFEARVNEQLYMMI	228
	NOV41b	169	NALGIGLSLYLICQVKAFLGCDVNLQNLFFGSLISAVDPVAVLAVFEARVNEQLYMMI	228
	gi 127814	169	NALGIGLSLYLICQVKAFLGCDVNLQNLFFGSLISAVDPVAVLAVFEARVNEQLYMMI	228
	gi 1346658	181	NSIGIGLSLFGICQIEAFGLSDITLLQNLFFGSLISAVDPVAVLAVFENIHVNEQLYILV	240
	gi 1709222	180	NSIGIGLSLFGICQIEAFGLSDITLLQNLFFGSLISAVDPVAVLAVFENIHVNEQLYILV	239
15	gi 15529998	180	NSIGIGLSLFGICQIEAFGLSDITLLQNLFFGSLISAVDPVAVLAVFENIHVNEQLYILV	239
	gi 6981560	65	NSIGIGLSLFGICQIEAFGLSDITLLQNLFFGSLISAVDPVAVLAVFENIHVNEQLYILV	124
<div>250260270280290300</div>				
20	NOV41a	229	FGEALLNDGITVVVLYNMLIAFTKMKHFEDTETVDILACCARFIIVGLGGVLFGIIVFGFI	288
	NOV41b	229	FGEALLNDGITVVVLYNMLIAFTKMKHFEDTETVDILACCARFIIVGLGGVLFGIIVFGFI	287
	gi 127814	229	FGEALLNDGITVVVLYNMLIAFTKMKHFEDTETVDILACCARFIIVGLGGVLFGIIVFGFI	287
	gi 1346658	241	FGESLLNDAVTVVVLYNLFKSFCEOMK---TITQVDVFAGIANFFVVGIGGVILGIFLGF	296
	gi 1709222	240	FGESLLNDAVTVVVLYNLFKSFCEOMK---TITQVDVFAGIANFFVVGIGGVILGIFLGF	295
25	gi 15529998	240	FGESLLNDAVTVVVLYNLFKSFCEOMK---TITQVDVFAGIANFFVVGIGGVILGIFLGF	295
	gi 6981560	125	FGESLLNDAVTVVVLYNLFKSFCEOMK---TITQVDVFAGIANFFVVGIGGVILGIFLGF	180
<div>310320330340350360</div>				
30	NOV41a	289	SAFITRFTONISATEPLIVFMFSYLSYLAETLYLSGILAITACAVTMKKYVEENVSC	348
	NOV41b	288	SAFITRFTONISATEPLIVFMFSYLSYLAETLYLSGILAITACAVTMKKYVEENVSC	347
	gi 127814	288	SAFITRFTONISATEPLIVFMFSYLSYLAETLYLSGILAITACAVTMKKYVEENVSC	347
	gi 1346658	297	AAFTTRFTNIRVIEPLFVFLSYLSYITABMFHLSGIMAITACAMTMNKYVEENVSC	356
	gi 1709222	296	AAFTTRFTNIRVIEPLFVFLSYLSYITABMFHLSGIMAITACAMTMNKYVEENVSC	355
35	gi 15529998	296	AAFTTRFTNIRVIEPLFVFLSYLSYITABMFHLSGIMAITACAMTMNKYVEENVSC	355
	gi 6981560	181	AAFTTRFTNIRVIEPLFVFLSYLSYITABMFHLSGIMAITACAMTMNKYVEENVSC	240
<div>370380390400410420</div>				
40	NOV41a	349	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFICFTLAFCLIWRAISVFAFY	408
	NOV41b	348	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFICFTLAFCLIWRAISVFAFY	407
	gi 127814	348	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFICFTLAFCLIWRAISVFAFY	407
	gi 1346658	357	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFVCFCLIAWRAISVFAFY	416
	gi 1709222	356	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFVCFCLIAWRAISVFAFY	415
45	gi 15529998	356	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFVCFCLIAWRAISVFAFY	415
	gi 6981560	241	YTTIKYFMKMLSSVSETLIFIFMGVSTVGKNHEWNWAFVCFCLIAWRAISVFAFY	300
<div>430440450460470480</div>				
50	NOV41a	409	SNQFRTEFFSFKDQCIIFYSQVIRGAGSESLAFLPLSLFPRKKMFVTATVVVIFTVF	468
	NOV41b	408	SNQFRTEFFSFKDQCIIFYSQVIRGAGSESLAFLPLSLFPRKKMFVTATVVVIFTVF	467
	gi 127814	408	SNQFRTEFFSFKDQCIIFYSQVIRGAGSESLAFLPLSLFPRKKMFVTATVVVIFTVF	467
	gi 1346658	417	INWFRTIPLTFKQFI IAYCGIRGAICFALVFLLEPAVFPKKLFITAAIVVIFTVF	476
	gi 1709222	416	INWFRTIPLTFKQFI IAYCGIRGAICFALVFLLEPAVFPKKLFITAAIVVIFTVF	475
55	gi 15529998	416	INWFRTIPLTFKQFI IAYCGIRGAICFALVFLLEPAVFPKKLFITAAIVVIFTVF	475
	gi 6981560	301	INWFRTIPLTFKQFI IAYCGIRGAICFALVFLLEPAVFPKKLFITAAIVVIFTVF	360
<div>490500510520530540</div>				
60	NOV41a	469	GITVGPLVRYLDVKKTNKKE-SINEEDHIRLMDHLKAGIEDVCGHWSHYQVRDKFKKFDH	527
	NOV41b	468	GITVGPLVRYLDVKKTNKKE-SINEEDHIRLMDHLKAGIEDVCGHWSHYQVRDKFKKFDH	526
	gi 127814	468	GITVGPLVRYLDVKKTNKKE-SINEEDHIRLMDHLKAGIEDVCGHWSHYQVRDKFKKFDH	526
	gi 1346658	477	GITIRPLVEFLDVKRSNKKQAVSEELHCRFFDHVKTGIEDVCGHWSHYQVRDKFKKFDH	536
	gi 1709222	476	GITIRPLVEFLDVKRSNKKQAVSEELHCRFFDHVKTGIEDVCGHWSHYQVRDKFKKFDH	535
65	gi 15529998	476	GITIRPLVEFLDVKRSNKKQAVSEELHCRFFDHVKTGIEDVCGHWSHYQVRDKFKKFDH	535
	gi 6981560	361	GITIRPLVEFLDVKRSNKKQAVSEELHCRFFDHVKTGIEDVCGHWSHYQVRDKFKKFDH	420
<div>550560570580590600</div>				
70	NOV41a	528	RYLRKILIRKNLPKSSIVSLYKKLEMKQAIEMVETGILSSSTAFSIPHOAQRIOGKRLSP	587
	NOV41b	527	RYLRKILIRKNLPKSSIVSLYKKLEMKQAIEMVETGILSSSTAFSIPHOAQRIOGKRLSP	586

5	gi 127814	527	KYLRLKLLIRENQPKSSIVSLYKKLEIKHAEMAETGMLISSVASTPYQSERIQGKRLSLP	586
	gi 1346658	537	KYLRLKLLIRENQPKSSIVSLYKKLEIKHAEMAETGMISTVPSFASLNDCREEKIRKLTP	596
	gi 1709222	536	KYLRLKLLIRENQPKSSIVSLYKKLEIKHAEMAETGMISTVPSFASLNDCREEKIRKLTP	595
	gi 15529998	536	KYLRLKLLIRENQPKSSIVSLYKKLEIKHAEMAETGMISTVPTFASLNDCREEKIRKVTS	595
	gi 6981560	421	KYLRLKLLIRENQPKSSIVSLYKKLEIKHAEMAETGMISTVPSFASLNDCREEKIRKLTP	480
10	NOV41a	588	EDVESIRDLILSNMYQVRORTLSYNKYNLKPOTSEKQAKEILIRRONTLRESMRKGHSLP	647
	NOV41b	587	EDVESIRDLILSNMYQVRORTLSYNKYNLKPOTSEKQAKEILIRRONTLRESMRKGHSLP	646
	gi 127814	587	EDVESIRDLILSNMYQVRORTLSYNKYNLKPOTSEKQAKEILIRRONTLRESMRKGHSLP	646
	gi 1346658	597	GEMDEIREILSRNLYQIRORTLSYNRHNLTADTSEKQAKEILIRRRHSLRESIRKDNLSL	656
	gi 1709222	596	GEMDEIREILSRNLYQIRORTLSYNRHNLTADTSEKQAKEILIRRRHSLRESIRKDNLSL	655
15	gi 15529998	596	SETDEIREILSRNLYQIRORTLSYNRHNLTADTSEKQAKEILIRRRHSLRESIRKDNLSL	655
	gi 6981560	481	GEMDEIREILSRNLYQIRORTLSYNRHNLTADTSEKQAKEILIRRRHSLRESIRKDNLSL	540
20	NOV41a	648	WGKPAGTKNIRYLSYPYGN--PQSACRDTRAAGFSG-----KLPTWL	687
	NOV41b	647	WGKPAGTKNIRYLSYPYGN--PQSACRDTRAAGFSG-----KLPTWL	686
	gi 127814	647	WGKPAGTKNIRYLSYPYGN--PQSACRDTRAAGFSG-----KLPTWL	686
	gi 1346658	657	REHRASTSTSRYSLSLPKNTKLPEKLQKKNKVSNADGNSSDSDMDG-TTVNLQPRARRFL	715
	gi 1709222	656	REHRASTSTSRYSLSLPKNTKLPEKLQKKNKVSNADGNSSDSEADAGTTVLNLQPRARRFL	715
25	gi 15529998	656	REHRASTSTSRYSLSLPKNTKLPEKLQKKRTTISADGNSSDSDADAGTTVLNLQPRARRFL	715
	gi 6981560	541	REHRASTSTSRYSLSLPKNTKLPEKLQKKNKVSNADGNSSDSDMDG-TTVNLQPRARRFL	599
30	NOV41a	688	LCCFSEVESGG-----KYLGVWAKROH-----	708
	NOV41b	687	LLWLRFGRGGLTMDTACTITG-----FIVLCSKKNSVIVHKIVLFLKSLSSYNL	737
	gi 127814	687	LHFLLCR--AMVE-----KIWGPGGQETOPRLL-----	712
	gi 1346658	716	PDQFSKKASPAYKMEWKNEVDVGSARAPESVTPAPRSKEGGTQTPGVLRQPLLSKDQRF	775
	gi 1709222	716	PEPFSKKASQAYKMEWKNEVDAGSGQGQSPPAAPRSKEGGTQTPAVLRQPLLSKDQ	772
35	gi 15529998	716	PEQFSKKSPQSYKMEWKNEVDVDSGRDMPSTPPTPHSREKGTQTSCLLQOPLLSKDQS-G	774
	gi 6981560	600	PDQFSKKASPAYKMEWKNEVDVGSARAPESVTPAPRSKEGGTQTPGVLRQPLLSKDQRF	659
40	NOV41a	708	ERRISGTFQKRAIEFAVLKLLNMDL-----	761
	NOV41b	738	ERRISGTFQKRAIEFAVLKLLNMDL-----	761
	gi 127814	713	CRNLN-----	717
	gi 1346658	776	RGREDSLITEDVPPKPPPLVRRASEPGNRKGRLGNEKP	813
	gi 1709222	772	-GREDSLITEGGRPKPPPLVRRASEPGNRKSRGSDKP	809
45	gi 15529998	775	SEREDSLITEGIPPKPPPLVRRASEPGSRKARFGSEKP	812
	gi 6981560	660	RGREDSLITEDGPPKPPPLVRRASEPGNRKGRLGNEKP	697

Na⁺/H⁺ antiporters are key transporters in maintaining the pH of actively metabolizing cells. Na⁺/H⁺ exchange proteins eject protons from cells, effectively eliminating excess acid from actively metabolising cells. Na⁺/H⁺ exchange activity is also crucial for the regulation of cell volume, and for the reabsorption of NaCl across renal, intestinal, and other epithelia. These antiports exchange Na⁺ for H⁺ in an electroneutral manner, and this activity is carried out by a family of Na⁺/H⁺ exchangers, or NHEs. In mammalian cells, Na⁺/H⁺ exchange activity is found in both the plasma membrane and inner mitochondrial membrane. To date, six mammalian isoforms have been identified (designated NHE1-NHE6). These exchangers are highly-regulated (glyco)phosphoproteins, which, based on their primary structure, appear to contain 10-12 transmembrane regions at the N-terminus and a large cytoplasmic region at the C-terminus. The transmembrane regions M3-M12 share identity with other members of the

family. The M6 and M7 regions are highly conserved. Thus, this is thought to be the region that is involved in the transport of sodium and hydrogen ions. The cytoplasmic region has little similarity throughout the family. There is some evidence that they may exist in the cell membrane as homodimers, but the molecular mechanisms of antiport are unclear. Na⁺/H⁺ antiporters play an important role in signal transduction.

The disclosed NOV41 nucleic acid of the invention encoding a Sodium/Hydrogen Exchanger 4 -like protein includes the nucleic acid whose sequence is provided in Table 41A, 41C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 41A or 41C while still encoding a protein that maintains its UDP[Glycosyltransferase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 16 percent of the bases may be so changed.

The disclosed NOV41 protein of the invention includes the Sodium/Hydrogen Exchanger 4 -like protein whose sequence is provided in Table 41B or 41D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 41B or 41D while still encoding a protein that maintains its Sodium/Hydrogen Exchanger 4 -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 37 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Sodium/Hydrogen Exchanger 4 -like protein (NOV41) is a member of a "Sodium/Hydrogen Exchanger 4 family". Therefore, the NOV41 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated

below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV41 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in brain disorders including hypercalcaemia, ulcers, inflammatory bowel disease, diverticular disease; diseases of the kidney including diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, and others; diseases of the brain including Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, epilepsy, and others; endometriosis, fertility, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, and/or other diseases and pathologies.

NOV41 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV41 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV41 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV42

NOV42 includes three novel Kupffer Cell Receptor -like proteins disclosed below. The disclosed sequences have been named NOV42a, NOV42b, NOV42c, and NOV42d.

NOV42a

A disclosed NOV42a nucleic acid of 1760 nucleotides (also referred to as CG56682-01) encoding a Kupffer Cell Receptor -like protein is shown in Table 42A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 16-18 and ending with a TGA codon at nucleotides 1661-1663. The start and stop codons are shown in bold in Table 42A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 42A. NOV42a nucleotide sequence (SEQ ID NO:157).

```
TGGCTGGGAGCAGTGCTGGAGGATGAAGGAAGCAGAGATGGACGGTGAGGCAGTCCGCTTCTGCACAGATAA
CCAGTGTGTCTCCCTGCACCCCAAGGTGTGGACTCTGTGGCAATGGCTCCTGCAGCCCCAAGATACCCGAG
GCTCGTTTCAGGCTACCCCGGCATTTATGGCTGTGACCTTGGTCTTCTCTCTTGTGACTCTCTTTGTAGTGGG
TAAGCCCCCAGGTGACCCAAATCTCACTAATTTCTCTCTTTCAGCACAAAGTCCCCAGGGGCCCCAGATG
CACACTCGATCATCACCATTGTCAGGGAGGCAGAAATGCGAGAGCTTATCCAGACATTTAAAGGCCACAT
GGAGAATTCAGTGCCTGGGTAGTAGAAATCCAGATGTTGAAGTGCAGAGTGGACAATGTCAATTCGCAGCT
CCAGGTGCTCGGTGATCATCTGGGAAACACCAATGCTGACATCCAGATGGTAAAAGGAGTTCTTAAAGGATGC
CACTACATTGAGTTTGCAGACACAGATGTTAAGGAGTTCCCTGGAGGGAACCAATGCTGAGATCCAGAGGCT
CAAGGAAGACCTTGAAAAGGCAGATGCTTTAACTTTCCAGACGCTGAATTTCTTAAAAAGCAGTTTAGAAAA
CACCAGCATTGAGCTCCACGTGCTAAGCAGAGGCTTAGAAAAATGCAAACTCTGAAATTCAGATGTTGAATGC
CAGTTTGGAAACGGCAAATGCTTTAACTCCCAGACCCAGGCCTTTATAAAAAGCAGTTTGTGACAACTAG
TGCTGAGATCCAGTTCTTAAAGAGGTCAATTTGGAAAGAGCTGGTGATGAAATTCACGTGTTAAAAAGGGATT
GAAAAATGGTCACAGCCCAGACCCAAAAAGCAATGGCCGCTCTGGACCAGACAGATACTCAGATTCAGGTATT
CAAGTCAGAGATGGAAAAATGTGAATACCTTAAATGCCAGATTCAGGTCTTAAATGGTCATATGAAAAATGC
CAGCAGAGAGATACAGACCCTAAACAAGGAATGAAGAATGCTTCAGCCTTAACTTCCAGACCCAGATGTT
AGACAGCAATCTGCAGAAGGCCAGTGCCGAGATCCAGAGGTTAAGAGGGGATCTAGAGAACACCAAAGCTCT
AACCATGGAAATCCAGCAGGAGCAGAGTCCCTGAAGACCTCCATGTGGTCATTACTTTCACAGGAACAGCT
ACAAAGAACCCAAAGTAAGCAGCTTCTCCAGATGGTCTGCAAGGCTGGAAGTTCAATGGTGGAAAGCTTATA
TTATTTTCTAGTGTCAAGAAGTCTTGGCATGAGGCTGAGCAGTTCTGCGTGTCCAGGGAGCCCATCTGGC
ATCTGTGGCCTCCAAGGAGGAGCAGGCATTTCTGGTAGAGTTCACAAGTAAAGTGTAAGTACTGATCGGTCT
CACTGACAGGGGCACAGAGGGCTCCTGGCGCTGGACAGATGGGACACCATTCAACGCCGCCAGAACAAAGG
GTTTTGGGAAAAGAAATCAGTCTGACAACTGGCGGCACAAGAATGGGCAGACTGAAGACTGTGTCCAAATTCA
GCAGAAGTGGAATGACATGACCTGTGACACCCCTATCAGTGGGTGTGCAAGAAGCCCATGGGCCAGGGTGT
GGCCTGAGGGGAGGCCAGAGCTGAGGGGCTGCTCCTGCTTGCCTAATCTGACCCCTCCTCCTGATGCCTTCG
GAGCCTCTGAGCTCTGCTTGTCTCTGGGACC
```

In a search of public sequence databases, the NOV42a nucleic acid sequence, located on chromosome 2, has 1214 of 1730 bases (70%) identical to a gb:GENBANK-ID:D88577|acc:D88577.1 mRNA from *Mus musculus* (mRNA for Kupffer cell receptor, complete cds) ($E = 3.9e^{-162}$).

The disclosed NOV42a polypeptide (SEQ ID NO:158) encoded by SEQ ID NO:157 has 546 amino acid residues and is presented in Table 42B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42a has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.7900. Alternatively, NOV42a may also localize to the microbody (peroxisome) with a certainty of 0.3000, to the Golgi body with a certainty of 0.3000, or to the endoplasmic reticulum (membrane) with a certainty of 0.2000. The most likely cleavage site for NOV42a is between positions 65 and 66: VVG-KP.

Table 42B. Encoded NOV42A protein sequence (SEQ ID NO:158).

```
MKEAEMDGEAVRFCTDNQCVSLHPQGVDSVAMAPAPKIPRLVQATPAFMAVTLVFSVLTLFVVGKPPGDPN
LTNFLSFQHKVPRGPRCTLDHHHFGREAEMRELIQTFKGHMENSASVVEIQMLKCRVDNVNSQLQVLGDHL
GNTNADIQMVKGVLKDATTLSLQTMRLRSSLEGTNAEIQRLKEDLEKADALTFQTLNFKSLSENTSIELHV
LSRGLNANSEIQMLNASLETANALNSQTQAFIKSSFNTSAEIQFLRGHLERAGDEIHVLKRDLMVTAQT
QKANGRLDQTDQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSNLQKA
SAEIQLRLRGDLENTKALTMETIQEQSRLKTLHVVTISQEQQLQRTQSKQLLQMVLLQGWKFNGGSLYFSSVKK
SWHEAEQFCVSQGAHLASVASKEEQAFLEFTSKVYYWIGLTDRTGTEGSRWTDGTFPNAQNKGFWEKNQS
DNWRHKNQGTEDCVQIQQKWNMTCDTPYQWVCKKPMGGQVA
```


A search of sequence databases reveals that the NOV42a amino acid sequence has 301 of 546 amino acid residues (55%) identical to, and 396 of 546 amino acid residues (72%) similar to, the 548 amino acid residue ptnr:SWISSPROT-ACC:P70194 protein from *Mus musculus* (Mouse) (Kupffer Cell Receptor) (E = 0.0).

NOV42a is predicted to be expressed in at least cartilage. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in :Kupffer cells (liver) because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:D88577|acc:D88577.1) a closely related *Mus musculus* mRNA for Kupffer cell receptor, complete cds homolog.

NOV42b

In the present invention, the target sequence identified previously, NOV42a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide

the sequence reported below, which is designated NOV42b. This differs from the previously identified sequence (NOV42a) in having 3 less aminoacids and 26 different ones.

A disclosed NOV42b nucleic acid of 1769 nucleotides (also referred to as CG56682-02) encoding a Kupffer cell receptor-like protein is shown in Table 42C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 23-25 and ending with a TGA codon at nucleotides 1670-1672. The start and stop codons are shown in bold in Table 42C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 42C. NOV42b nucleotide sequence (SEQ ID NO:159).

TGGCTGGGAGCAGTGTCTGGAGGATGAAGGAAGCAGAGATGGACGGTGAGGCAGTCCGCTTCTGCACAGATAA
 CCAGTGTGTCTCCCTGCACCCCCAAGAGGTGGACTCTGTGGCAATGGCTCCTGCAGCCCCAAGATACCGAG
 GCTCGTTTCAGGCTACCCCGGCATTTATGGCTGTGACCTTGGTCTTCTCTTGTGACTCTCTTTGTAGTGGT
 TCAACAGCAGACAAGACCTGTTCCGAAGCCTGTGCAAGCCGTAATTCTGGGAGACAACATTACTGGGCATTT
 ACCTTTTGAACCCAACAATCATCACCACTTTGGCAGGGAGGCAGAAATGCGAGAGCTTATCCAGACATTTAA
 AGGTCACATGGAGAATTCCAGTGCCTGGGTAGTAGAAATCCAGATGTTGAAGTGCAGAGTGGACAATGTCAA
 TTCGACAGTCCAGGTGCTCGGTGATCATCTGGGAAACACCAATGCTGACATCCAGATGGTAAAGGAGTTCT
 AAAGGATGCCACTACATTGAGTTTGCAGACACAGATGTTAAGGAGTTCCTTGAGGGAACCAATGTCTGAGT
 CCAGAGGCTCAAGGAAGACCTTGAAGGAGCAGATGCTTTAACTTTCCAGACGCTGAATTTCTTAAAAAGCAG
 TTTAGAAAACACCAGCATTGAGCTCCACGTGCTAAGCAGAGGCTTAGAAAATGCAAACCTCTGAAATTCAGAT
 GTTGAATGCCAGTTTGGAAACGGCAAATGCTTTAACTCCCAGACCCAGGCCTTTATAAAAAGCAGTTTGA
 CAACACTAGTCTGAGATCCAGTTCTTAAGAGGTCATTTGGAAAGAGCTGGTGATGAAATTCACGTGTTAAA
 AAGGGATTTGAAATGGTCACAGCCCAGACCCAAAAAGCAAATGGCCGTCTGGACCAGACAGATACTCAGAT
 TCAGGTATTCAAGTCAGAGATGGAAATGTGAATACCTTAAATGCCAGATTCAGGTCTTAAATGGTCATAT
 GAAAAATGCCAGCAGAGAGATACAGACCCCTAAAAACAAGGAATGAAGAATGCTTCAGCCTTAACCTCCAGAC
 CCAGATGTTAGACAGCAATCTGCAGAAGGCCAGTGCCGAGATCCAGAGGTTAAGAGGGATCTAGAGAACAC
 CAAAGCTCTAACCATGGAAATCCAGCAGGAGCAGAGTCGCTGAAGACCCTCCATGTGGTCATTACTTCACA
 GGAACAGCTACAAAAGAACCCAAAGTAAGCAGCTTCTCCAGATGGTCTGCAAGGCTGGAAGTTCAATGGTGG
 AAGCTTATAATATTTTCTAGTGTCAAGAAGTCTTGGCATGAGGCTGAGCAGTTCTGCGTGTCCAGGGAGC
 CCATCTGGCATCTGTGGCTCCAAGGAGGAGCAGGCATTTCTGGTAGAGTTACAAAGTAAAGTGTAAGTCTG
 GATCGGTCTCACTGACAGGGGCACAGAGGGCTCCTGGCGCTGGACAGATGGGACACCATTCAACGCCGCCCA
 GAACAAAGGGTTTGGGAAAAGAATCAGTCTGACAACCTGGCGGCACAAGAATGGGCAGACTGAAGACTGTGT
 CCAAATTCAGCAGAAGTGAATGACATGACCTGTGACACCCCTATCAGTGGGTGTGCAAGAAGCCCATGGG
 CCAGGGTGTGGCCTGAGGGCAGGCCAGAGCTGAGGGGCTGCTCCTGCTTGCCAATACTGACCTCCTCCTCG
 ATGCCCTCGGAGCCTCTGAGCTCTGCTTGTCTCTGGGACC

In a search of public sequence databases, the NOV42b nucleic acid sequence, located on chromosome 2, has 1054 of 1469 bases (71%) identical to a gb:GENBANK-ID:D88577|acc:D88577.1 mRNA from *Mus musculus* (mRNA for Kupffer cell receptor, complete cds) ($E = 1.1e^{-161}$).

The disclosed NOV42b polypeptide (SEQ ID NO:160) encoded by SEQ ID NO:159 has 549 amino acid residues and is presented in Table 42D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42b has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.7900. Alternatively, NOV42b may also localize to the microbody (peroxisome) with a certainty of 0.3000, to the Golgi body with a certainty of 0.3000, or to the endoplasmic reticulum (membrane) with a certainty of 0.2000. The most likely cleavage site for NOV42b is between positions 67 and 68: VQQ-QT.

Table 42D. Encoded NOV42b protein sequence (SEQ ID NO:160).

MKEAEMDGEAVRFCTDNQCVSLHPQEVDSVAMAPAAPKIPRLVQATPAFMAVTLVFSLVTLFVVVQQQTRPV
PKPVQAVILGDNITGHLPPFEPNNHHHFGREAEMLRIQTFKGHMENSASWVVEIQMLKCRVDNVNSQLQVLG
DHLGNTNADIQMVKGVLKDATTLSLQTMRLSSLEGTNAEIQRLEDLEKADALTFQTLNFKSSLENTSIE
LHVLSRGLNANSEIQMLNASLETANALNSQTQAFIKSSFNDNTSAEIQFLRGHLERAGDEIHVLKRDLMVMT
AQTKANGRLDQTDQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSNL
QKASAEIQRLRGDLNLTALMEIQEQSRLKTLHVVITSQEQLQRTQSKQLLQMVLLQGWKFNGGSLYFSS
VKKSWHEAEQFCVSQGAHLASVASKEEQAFLEFTSKVYWI GLTDRGTEGSRWTDGTPFNAAQNKGFWK
NQSDNWRHKNQGTEDCVQIQKWNMTCDTPYQVWCKKPMQGVV

A search of sequence databases reveals that the NOV42b amino acid sequence has 304
of 549 amino acid residues (55%) identical to, and 401 of 549 amino acid residues (73%)
similar to, the 548 amino acid residue ptnr:SWISSPROT-ACC:P70194 protein from *Mus*
musculus (Mouse) (Kupffer Cell Receptor) ($E = 3.7e^{-158}$).

NOV42b is predicted to be expressed in at least adrenal gland, bone marrow, brain -
amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus,
brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji,
mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle,
small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

NOV42c

A disclosed NOV42c nucleic acid of 1874 nucleotides (also referred to as CG56682-
03) encoding a Kupffer cell receptor-like protein is shown in Table 42E. An open reading
frame was identified beginning with a ATG initiation codon at nucleotides 1-3 and ending
with a TAA codon at nucleotides 1702-1704. The start and stop codons are shown in bold in
Table 42E, and the 5' and 3' untranslated regions, if any, are underlined.

Table 42E. NOV42c nucleotide sequence (SEQ ID NO:161).

ATGGACGGTGAGGCAGTCCGCTTCTGCACAGATAAC**CAGTGTGTCTCC**TGCACCCCCAAGAGGTGGACTCT
GTGGCAATGGCTCCTGCAGCCCCCAAGATACCGAGGCTCGTTCAGGCTACCCCGGCATTATGGCTGTGACC
TTGGTCTTCTCTCTTGTGACTCTCTTGTAGTGGTTCAACAGCAGACAAGACCTGTTCCGAAGCCTGTGCAA
GCCGTAATTTCTGGGAGACAACATTACTGGGCATTTACCTTTTGAACCCAACAATCATCACCACCTTTGGCAGG
GAGGCAGAAATGCAAGAGCTTATCCAGACATTTAAAGGCCACATGGAGAATTCCAGTGCCTGGGTAGTAGAA
ATCCAGATGTTGAAGTGCAGAGTGGACAATGTCATTCGAGCTCCAGGTGCTCGGTGATCATCTGGGAAAC
ACCAATGCTGCATCCAGATGGTAAAGGAGTCTTAAAGGATGCCACTACATTGAGTTTGCAGACACAGATG
TTAAGGAGTTCCTGGAGGGAACCAATGCTGAGATCCAGAGGCTCAAGGAAGACCTTGAAGAGCAGATGCT
TTAACTTTCCAGACGCTGAATTTCTTAAAAAGCAGTTTAGAAAACACCAGCATTGAGCTCCAGTGCTAAGC
AGAGGCTTAGAAAATGCAAACTCTGAAATTCAGATGTTGAATGCCAGTTTGGAAACGGCAATACCCAGGCT
CAGTTAGCCAAATAGCAGTTTAAAGAACGCTAATGCTGAGATCTATGTTTGGAGAGCCATTAGATAGTGTC
AATGACTTGAGGACCCAGAACCAGGTTTAAAGAAATAGTTTGAAGGAGCCAATGCTGAGATCCAGGGACTA
AAGGAAAATTTGCAGAACACAAATGCTTTAACTCCCAGACCCAGGCCTTTATAAAAAGCAGTTTGTGACAAC
ACTAGTGCTGAGATCCAGTTCTTAAGAGGTCATTTGGAAAGAGCTGGTGATGAAATTCAGTGTATAAAGG
GATTTGAAAATGGTTCACAGCCAGACCCAAAAGCAAAATGGCCATCTGGACAGACAGATACTCAGATTTCAG
GTATTCAGTGCAGAGATGGAAAATGGAATACCTTAAATGCCAGATTCAGGTCTTAAATGGTCATATGAAA
AATGCCAGCAGAGAGATACAGACCCTAAAACAGGAATGAAGAATGCTTCAGCCTTAACTCCCAGACCCAG
ATGTTAGACAGCAATCTGCAGAAGGCCAGTGGCAGATCCAGAGGTTAAGAGGGGATCTAGAGAACACCAA
GCTCTAACCATGGAAATCCAGCAGGAGCAGAGTGCCTGAAGACCCTCCATGTGGTCATTACTTCACAGGAA
CAGCTACAAAGAACCCAAAGTCAGCTTCTCCAGATGGTCTGCAAGGCTGGAAGTTCAATGGTGAAGCTTA
TATTATTTTTCTAGTGTCAAGAAGCTTGGCATGAGGCTGAGCAGTTCTGCGTGTCCAGGGAGCCCATCTG
GCATCTGTGGCCTCCAAGGAGGAGCAGGCATTTCTGGTAGAGTTCACAAGTAAAGTGTACTACTGGATCGGT

CTCACTGACAGGGGCACAGAGGGCTCCTGGCGCTGGACAGATGGGACACCATTCAACGCCGCCAGAACAAA
 GCCTCCCTAGGAGCCACAGCACCAGGAAGGGATGCTGCCTTCATCTAACAGTATAAAGCCCTGTTGTCTTCG
 GGT'TTTGGGAAAAGAATCAGTCGACAACCTGGCGGCACAGAATGGGCAGACTGAAGACTGTGTCCAAATTCA
 GCAGAAGTGGAATGACATGACCTGTGACACCCCTATCAGTGGGTGTGCAAGAAGCCCATGGGCCAGGGTGT
 GG

In a search of public sequence databases, the NOV42c nucleic acid sequence, located on chromosome 2, has 689 of 993 bases (69%) identical to a gb:GENBANK-ID:D88577|acc:D88577.1 mRNA from *Mus musculus* (mRNA for Kupffer cell receptor, complete cds) ($E = 1.6e^{-120}$).

The disclosed NOV42c polypeptide (SEQ ID NO:162) encoded by SEQ ID NO:161 has 567 amino acid residues and is presented in Table 42F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42c has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.7900. Alternatively, NOV42c may also localize to the microbody (peroxisome) with a certainty of 0.3000, to the Golgi body with a certainty of 0.3000, or to the endoplasmic reticulum (membrane) with a certainty of 0.2000. The most likely cleavage site for NOV42c is between positions 62 and 63: VQQ-QT.

Table 42F Encoded NOV42c protein sequence (SEQ ID NO:162).

MDGEAVRFCTDNQCVS LHPQEVDSVAMAPAAPKIPRLVQATPAFMAVTLVFSLVTLFVVVQQQTRPVPKPVQ
 AVILGDNITGHLPPFEPNNHHHFGREAEQMELIQTFKGHMENS SAWVVEIQMLKCRVDNVNSQLQVLGDHLGN
 TNADIQMVKGVLKDATTLSLQTQMLRSSLEG TNAEIQRKEDLEKADALTFQTLNFLKSSLENTSIELHVL
 RGLNANSEIQMLNASLETANTQAQLANSS LKNANAEIYVLRGHLDSVNDLRTQNQVLRNSLEGANAEIQGL
 KENLQNTNALNSQTQAFIKSSF DNTSAEIQFLRGHLERAGDEIHVLKRD LKMVTAQTQKANGHLDQTD TQIQ
 VFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSNLQKASAEIQRLRGDLENTK
 ALTMEIQEQSRLKTLHVVITSQEQLQRTQS QLLQMV LQGKWFNGGSLYFSSVKKSWHEAEQFCVSQGAHL
 ASVASKEEQAF LVEFTSKVYYWIGLTD RGTGEGSWRWDGTPFNAAQNKASLGATAPGRDAAFI

A search of sequence databases reveals that the NOV42c amino acid sequence has 191 of 412 amino acid residues (46%) identical to, and 273 of 412 amino acid residues (66%) similar to, the 548 amino acid residue ptnr:SWISSNEW-ACC:P70194 protein from *Mus musculus* (Mouse) (Kupffer Cell Receptor) ($E = 5.3e^{-92}$).

NOV42c is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

NOV42d

A disclosed NOV42d nucleic acid of 1985 nucleotides (also referred to as CG56682-04) encoding a Kupffer cell receptor-like protein is shown in Table 42G. An open reading frame was identified beginning with a GTC initiation codon at nucleotides 2-4 and ending with a TAA codon at nucleotides 1658-1660. The start and stop codons are shown in bold in Table 42G, and the 5' and 3' untranslated regions, if any, are underlined. Because the starting codon is not a traditional initiation codon, NOV42d could be a partial reading frame extending further into the 5'

Table 42G. NOV42d nucleotide sequence (SEQ ID NO:163).

AGTCCGCTTCTGCACAGATAACCAAGTGTGTCTCCCTGCACCCCAAGAGGTGGACTCTGTGGCAATGGCTCC
TGCAGCCCCCAAGATACCGAGGCTCGTTCAGGCTACCCCGGCATTTATGGCTGTGACCTTGGTCTTCTCTCT
TGTGACTCTCTTGTAGTGGTTCAACAGCAGACAAGACCTGTTCCGAAGCCTGTGCAAGCCGTAATTC'TGGG
AGACAACATTACTGGGCATTTACCTTTTGAACCAACAATCATCACCACCTTTGGCAGGGAGGCAGAAATGCA
AGAGCTTATCCAGACATTTAAAGGCCACATGGAGAATTCAGTGCCTGGGTAGTAGAAATCCAGATGTTGAA
GTGCAGAGTGGACAATGTCAATTCGCAGCTCCAGGTGCTCGGTGATCATCTGGGAACACCAATGCTGACAT
CCAGATGGTAAAGGAGTTCTAAAGGATGCCACTACATTGAGTTTGCAGACACAGATGTTAAGGAGTTCCT
GGAGGGAACCAATGCTGAGATCCAGAGGCTCAAGGAAGACCTTGAAAAGGCAGATGCTTTAACTTTCCAGAC
GCTGAATTTCTTAAAAAGCAGTTTAGAAAAACACCAGCATTGAGCTCCACGTGCTAAGCAGAGGCTTAGAAAA
TGCAAACTCTGAAATTCAGATGTTGAATGCCAGTTTGGAAACGGCAAATACCCAGGCTCAGTTAGCCAATAG
CAGTTTAAAGAACGCTAATGCTGAGATCTATGTTTTGAGAGGCCATCTAGATAGTGTCAATGACTTGAGGAC
CCAGAACCAGGTTTAAAGAAATAGTTTGGAAAGGAGCCAATGCTGAGATCCAGGGACTAAAGGAAAATTTGCA
GAACACAAATGCTTTAACTCCCAGACCCAGGCCTTTATAAAAAGCAGTTTGGCAACACTAGTGCTGAGAT
CCAGTTCTTAAAGAGGTCATTTGGAAAGAGCTGGTGATGAAATTCACGTGTTAAAAAGGGATTGAAAATGCT
CACAGCCCAGACCCAAAAAGCAAATGGCCGTCTGGACCAGACAGATACTCAGATTCAGGTATTCAAGTCAGA
GATGGAATATGTAATACCTTAAATGCCAGATTCAGGTCTTAAATGGTCATATGAAAAATGCCAGCAGAGA
GATACAGACCCTAAAAACAAGGAATGAAGAATGCTTCAGCCTTAACTCCCAGACCCAGATGTTAGACAGCAA
TCTGCAGAAAGGCCAGTGCCGAGATCCAGAGGTTAAGAGGGGATCTAGAGAACACCAAAGCTCTAACCATGGA
AATCCAGCAGGAGCAGAGTCGCCTGAAGACCCTCCATGTGGTCATTACTTCACAGGAACAGCTACAAAGAAC
CCAAAGTCAGCTTCTCCAGATGGTCTGCAAGGCTGGAAGTTCAATGGTGAAGCTTATATTATTTTCTAG
TGTCAAGAAGTCTTGGCATGAGGCTGAGCAGTTCTGCGTGTCCAGGGAGCCCATCTGGCATCTGTGGCCTC
CAAGGAGGAGCAGGCATTTCTGGTAGAGTTCACAAGTAAAGTGTACTACTGGATCGGTCTCACTGACAGGGG
CACAGAGGGCTCCTGGCGCTGGACAGATGGGACACCATTCAACGCCGCCAGAACAAAGCGGCCACTAGGGG
ATGAAGGACCCATCTCAAGTCAGCTCCCTAGACTCATCCCATGTTCAGCTCCCTAGGAGCCACAGCACCAGGA
AGGGATGCTGCTTCACTTAACAGTATAAAGCCCTGTTGTCTTCGGGTTTGGGAAAGAAATCAGTCTGACA
ACTGGCGGCACAAGAATGGGCAGACTGAAGACTGTGTCCAAATTCAGCAGAAGTGAATGACATGACCTGTG
ACACCCCTATCAGTGGGTGTGCAAGAAGCCATGGGCCAGGGTGTGGCCTGAGGGCAGGCCAGAGCTGAGG
GGTGCTCCTGCTTGCCAACTACTGACCTCCTCTCGATGC

In a search of public sequence databases, the NOV42d nucleic acid sequence, located on chromosome 2, has 705 of 1023 bases (68%) identical to a gb:GENBANK-ID:D88577|acc:D88577.1 mRNA from *Mus musculus* (mRNA for Kupffer cell receptor, complete cds) ($E = 3.7e^{-124}$).

The disclosed NOV42d polypeptide (SEQ ID NO:164) encoded by SEQ ID NO:163 has 552 amino acid residues and is presented in Table 42H using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV42d has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.7900. Alternatively, NOV42d may also localize to the microbody (peroxisome) with a certainty of 0.3000, to the Golgi body with a certainty of 0.3000, or to the endoplasmic reticulum (membrane) with a

certainty of 0.2000. The most likely cleavage site for NOV42d is between positions 57 and 58: VQQ-QT.

Table 42H. Encoded NOV42d protein sequence (SEQ ID NO:164).

```
VRFCTDNQCVSILHPQEVDSVAMAPAAPKIPRLVQATPAFMAVTLVFSLVTLFVVVQQQTRPVPKPVQAVILG
DNITGHLPPFEPNNHHHFGREAEQMELIQTFKGHMENS SAWVVEIQMLKCRVDNVNSQLQVLGDHLGNTNADI
QMVKGVLKDATLTLSTQMLRSSLEGTNAEIQRLKEDLEKADALTFQTLNFLKSSLENTSIELHVLSRGLEN
ANSEIQMLNASLETANTQAQLANSSLKANAEIYVLRGHLDSVNDLRTQNQVLRNSLEGANAEIQGLKENLQ
NTNALNSQTQAFIKSSFNTSAEIQFLRGHLERAGDEIHVLKRDLMVTAQTQKANGRLDQTDQIQVFKSE
MENVNTLNAQIQVLNGHMKNASREIQTQKGMKNASALTSQTQMLDSNLQKASAEIQRLRGDLENTKALTME
IQQEQSRLKTLHVVITSQEQLQRTQSLLQMVLGWKFNGGSLYYFSSVKKSWHEAEQFCVSSQGAHLASVAS
KEEQAFLEFTSKVYYWIGLTDRTGTEGSRWTDGTPFNAAQNKAATRG
```

A search of sequence databases reveals that the NOV42d amino acid sequence has 187 of 404 amino acid residues (46%) identical to, and 269 of 404 amino acid residues (66%) similar to, the 548 amino acid residue ptnr:SWISSNEW-ACC:P70194 protein from *Mus musculus* (Mouse) (Kupffer Cell Receptor) ($E = 1.2e^{-89}$).

NOV42d is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. .

NOV42a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 42I.

Table 42I. BLAST results for NOV42a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7949066 ref NP_058031.1 (NM_016751)	C-type (calcium dependent, carbohydrate recognition domain) lectin, superfamily member 13; kupffer cell receptor; Kupffer cell c-type lectin receptor [<i>Mus musculus</i>]	548	300/548 (54%)	395/548 (71%)	e-153
gi 16758588 ref NP_446205.1 (NM_053753)	Kupffer cell receptor [<i>Rattus norvegicus</i>]	550	293/547 (53%)	382/547 (69%)	e-147
gi 7657291 ref NP_056532.1 (NM_015717)	Langerhans cell specific c-type lectin; langerin [<i>Homo sapiens</i>]	328	92/261 (35%)	143/261 (54%)	2e-41
gi 17426713 emb CAC85632.1 (AJ313164)	langerin [<i>Mus musculus</i>]	326	91/264 (34%)	140/264 (52%)	1e-40

5

10

15

30

40

50

5	gi 7949066	116	SSSCHKEVQLKYQMDNVSSSLVOLLGSHLEVDNADILQTKDVLKESGALALETQALRSSI	175
	gi 16758588	116	SSSCHKEIQLKYQMDNVSSQVOLLGGHLEENADIQQAKDVLKGTGALASETQALRSSI	175
	gi 7657291	69	-----TISDVKTINVOLLKGRVDNIS	89
	gi 17426713	67	-----KILDVKSDAQMLKGRVDNIS	87
	gi 17059581	72	-----KILDVKSDAQMLKGRVDNIS	92
<div>190200210220230240</div>				
10	NOV42a	176	EGTNAEIQRLEKEDLEKADALTFTQLNFKSSLENTSIELHVLSRGLNANSEIQMLNASL	235
	NOV42b	179	EGTNAEIQRLEKEDLEKADALTFTQLNFKSSLENTSIELHVLSRGLNANSEIQMLNASL	238
	NOV42c	174	EGTNAEIQRLEKEDLEKADALTFTQLNFKSSLENTSIELHVLSRGLNANSEIQMLNASL	233
	NOV42d	169	EGTNAEIQRLEKEDLEKADALTFTQLNFKSSLENTSIELHVLSRGLNANSEIQMLNASL	228
15	gi 7949066	176	EVASADIHSIRGDLKANAMTSQTRGILLKSSSTENTSIAELHVLGRGLEBAQSEIQALRGSL	235
	gi 16758588	176	EVASADIHSIRGDLKANAMTSQTOGLLKSSSTENTSIAELHVLGRGLEBAQSEIQALRGSL	235
	gi 7657291	89	-----TIDSEIKKN-----SDGMEAGVQIQMVNESIGYVRSQFLKLKTSV	130
	gi 17426713	87	-----TIGSDLKTE-----RGRVDDAEVQMQIVNTTLKRVRSQILSLETSM	128
	gi 17059581	92	-----TIGSDLKTE-----RGRVDDAEVQMQIVNTTLKRVRSQILSLETSM	133
<div>250260270280290300</div>				
20	NOV42a	236	ETAN-----	239
	NOV42b	239	ETAN-----	242
	NOV42c	234	ETANTQAQLANSSLKNNANAEIYVLRGHLDSVNDLRTQNQVLRNSLEGANAEIQGLKENLQ	293
	NOV42d	229	ETANTQAQLANSSLKNNANAEIYVLRGHLDSVNDLRTQNQVLRNSLEGANAEIQGLKENLQ	288
25	gi 7949066	236	QSAN-----	239
	gi 16758588	236	QSSN-----	239
	gi 7657291	131	EKAN-----	134
	gi 17426713	129	KIAN-----	132
	gi 17059581	134	KIAN-----	137
<div>310320330340350360</div>				
35	NOV42a	239	---ALNSQTCAFIKSSFNTSAEIQFLRGHLERAGDETHVLRDLKMTAQTQKANGRLD	296
	NOV42b	242	---ALNSQTCAFIKSSFNTSAEIQFLRGHLERAGDETHVLRDLKMTAQTQKANGRLD	299
	NOV42c	294	NTNALNSQTCAFIKSSFNTSAEIQFLRGHLERAGDETHVLRDLKMTAQTQKANGRLD	353
	NOV42d	289	NTNALNSQTCAFIKSSFNTSAEIQFLRGHLERAGDETHVLRDLKMTAQTQKANGRLD	348
40	gi 7949066	239	---DLSSQTCGELQHSMDNISAEIQTVRDGMRAGEKMNLSLKELETTLTAQTQKANGHLE	296
	gi 16758588	239	---DLGSRTQNLQHSMDNISAEIQAMRDGMORAGEEMTSLLKKDLETTLTAQIQNANGHLE	296
	gi 7657291	134	-----	134
	gi 17426713	132	-----	132
	gi 17059581	137	-----	137
<div>370380390400410420</div>				
45	NOV42a	297	QTDITQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSN	356
	NOV42b	300	QTDITQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSN	359
	NOV42c	354	QTDITQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSN	413
	NOV42d	349	QTDITQIQVFKSEMENVNTLNAQIQVLNGHMKNASREIQTLKQGMKNASALTSQTQMLDSN	408
50	gi 7949066	297	QTDITQIQGLKAEKSTSSLSNRIEVNGOMKQDASRELOTLRRDLSDVSALKSNVQMLQSN	356
	gi 16758588	297	QTDITQIQGLKAEKSTSSLSNRIEVNGKLDSSRELOTLRRDLSDVSALKSNVQMLQSN	356
	gi 7657291	134	---AQIQILTRSWEEVSTLN-----	151
	gi 17426713	132	---DQIQILTRSWGEVDSLS-----	149
	gi 17059581	137	---DQIQILTRSWGEVDSLS-----	154
<div>430440450460470480</div>				
55	NOV42a	357	LQKASAEIQRLRGDLNLTALTMETIQEBSRLKTLHVVTTSQEBOLORTOSKOLLQMVLOG	416
	NOV42b	360	LQKASAEIQRLRGDLNLTALTMETIQEBSRLKTLHVVTTSQEBOLORTOSKOLLQMVLOG	419
	NOV42c	414	LQKASAEIQRLRGDLNLTALTMETIQEBSRLKTLHVVTTSQEBOLORTOS-OLLQMVLOG	472
	NOV42d	409	LQKASAEIQRLRGDLNLTALTMETIQEBSRLKTLHVVTTSQEBOLORTOS-OLLQMVLOG	467
60	gi 7949066	357	LQKAKAEVQSLKTGLEATKTLAAKIQGQSDLEALOKAVAAHTCGOKTON-OVLQIMQD	415
	gi 16758588	357	LQKAKAEVQSLKTGLEATKTLAAKIQGQSDLEALOKAVAAHTCGOKTON-OVLQIMQD	415
	gi 7657291	151	---AQIPELRSDLKASALNTKIKRALQGSLENMSKLLKRON-----DILEMVARG	198
	gi 17426713	149	---AKIPELKRDLDKASALNTKVOGLONSLENVKNLLKQCS-----DILEMVARG	196
	gi 17059581	154	---AKIPELKRDLDKASALNTKVOGLONSLENVKNLLKQCS-----DILEMVARG	201
<div>490500510520530540</div>				
70	NOV42a	417	WKENGCSLYYFSSVKKSWHEAEQFCVSQGAHLASVASKEEQAFLEFVTSKVVYIWIGLTDR	476
	NOV42b	420	WKENGCSLYYFSSVKKSWHEAEQFCVSQGAHLASVASKEEQAFLEFVTSKVVYIWIGLTDR	479

gnl|Pfam|pfam00059, lectin_c, Lectin C-type domain. This family includes both long and short form C-type (SEQ ID NO:837)
CD-Length = 107 residues, 99.1% aligned
Score = 115 bits (288), Expect = 6e-27

5

gnl|Pfam|pfam01576, Myosin_tail, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament. (SEQ ID NO:838)
CD-Length = 860 residues, 29.4% aligned
Score = 43.1 bits (100), Expect = 4e-05

30

Table 42N Domain Analysis of NOV42c

gnl|Pfam|pfam01576, Myosin_tail, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament. (SEQ ID NO:838)
CD-Length = 860 residues, 30.2% aligned
Score = 39.3 bits (90), Expect = 6e-04

5	NOV42:	135	SQLQVLGDHLGNTNADIQMVKGVLKDATTLSLQTQMLRSSLEGTNAEIQRLEKEDLEKADA	194
	Sbjct:	556	NELEIALDHANKANAEAQ-----KNVKYQQQVKELQTQVE---EEQRAREDAREQLA	605
	NOV42:	195	LTFQTLNFKSSLENTSIELHVLRSGLNANSEIQMLNASLETANALNSQTQAFIKSSFD	254
	Sbjct:	606	VAERRATALEAELEELRSALEQAERARKQAETE---LAEASERVNELTAQNSSLIAQK-R	661
10	NOV42:	255	NTSAEIQFLRGHLERAGDEIHVLKRDLMVTAQTQKANGRLDQTDQIQVFKSEMENVNT	314
	Sbjct:	662	KLEGELAALQSDLDEAVNELKAAEE-----RAKKAQADAARLAEELRQEHEHSQHLE	714
15	NOV42:	315	LNAQIQVLNGHMKNASREIQT--LKQGMKNASALTSQTQMLDSNL---QKASAEIQR-LR	368
	Sbjct:	715	LRKQLESQVKELQVRLDEAEAAALKGGKKMIQKLEARVRELEAELDGEQRRHAETQKNLR	774
20	NOV42:	369	GDLENTKALTMEIQQEQSRLKTLHVVITSQEQLQRTQSKQL	409
	Sbjct:	775	KMERRVKELQFQVEEDKKNLERLQDLVDKLQAKIKTYKRQL	815

Kupffer cells are found in the linings of the liver sinusoids, and are phagocytic. A receptor uniquely found on the surface of rat Kupffer cells binds oligosaccharides terminating in galactose, N-acetylgalactosamine, and fucose. A number of different families of proteins share a conserved domain which was first characterized in some animal lectins. Animal lectins display a wide variety of architectures. They are classified according to the carbohydrate-recognition domain (CRD) of which there are two main types, S-type and C-type. C-type lectins (CTL) display a wide range of specificities and function as a calcium-dependent carbohydrate-recognition domain. They are found predominantly but not exclusively in vertebrates. CTLs can be classified into a number of subgroups based on their function and structure: 1) Collectins, represented by the soluble mannose-binding proteins of mammalian serum and liver; 2) Selectins, membrane-bound proteins involved in inflammation; and 3) Endocytic lectins, membrane-bound receptors that mediate endocytosis of glycoproteins. Endocytic lectins are type-II membrane proteins where the CTL domain is located at the C-terminal extremity of the proteins, and include the Kupffer Cell Receptor.

The disclosed NOV42 nucleic acid of the invention encoding a Kupffer Cell Receptor-like protein includes the nucleic acid whose sequence is provided in Table 42A, 42C, 42E, 42G, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any

of whose bases may be changed from the corresponding base shown in Table 42A, 42C, 42E, or 42G while still encoding a protein that maintains its Kupffer Cell Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including
5 nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the
10 chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 32 percent of the bases may be so changed.

The disclosed NOV42 protein of the invention includes the Kupffer Cell Receptor -
15 like protein whose sequence is provided in Table 42B, 42D, 42F, or 42H. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 42B, 42D, 42F, or 42H while still encoding a protein that maintains its Kupffer Cell Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 66 percent of the
20 residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Kupffer Cell Receptor -like protein (NOV42) is a member of a "Kupffer Cell Receptor family". Therefore, the NOV42 nucleic
25 acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene
30 delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV42 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation, arthritis, tendinitis, and/or other diseases and pathologies.

NOV42 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV42 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV42 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV43

A disclosed NOV43 nucleic acid of 1108 nucleotides (also referred to as CG56690-01) encoding a P2Y Purinoceptor -like protein is shown in Table 43A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 12-14 and ending with a TAA codon at nucleotides 1095-1097. The start and stop codons are shown in bold in Table 43A; and the 5' and 3' untranslated regions, if any, are underlined.

Table 43A. NOV43 nucleotide sequence (SEQ ID NO:165).

```

GTCATGATGTTATGCTGTCCATTTTGCTTCCTTCCAGGGGAAGCAGAAGCGGGAGCCGTCGTGGAGCTCTGC
TCCTGGAGGGAGCCTCCCGGGACATGGAGAAGGTGGACATGAATACATCACAGGAACAAGGTCTCTGCCAGT
TCTCAGAGAAGTACAAGCAAGTCTACCTCTCCCTGGCCTACAGTATCATCTTTATCCTAGGGCTGCCACTAA
ATGGCACTGTCTTGTGGCACTCCTGGGGCCAAACCAAGCGCTGGAGCTGTGCCACCACCTATCTGGTGAACC
TGATGGTGGCCGACCTGCTTTATGTGCTATTGCCCTTCCTCATCATCACCTACTCACTAGATGACAGGTGGC
CCTTCGGGGAGCTGCTCTGCAAGCTGGTGCACCTCCTGTTCTATATCAACCTTTACGGCAGCATCCTGCTGC
TGACCTGCATCTCTGTGCACCAAGTTCCTAGGTGTGTGCCACCCACTGTGTTTCGCTGCCCTACCGGACCCGCA
GGCATGCCTGGCTGGGCACCAAGCACCACCTGGGCCCTGGTGGTCCCTCCAGCTGCTGCCACACTGGCCCTCT
CCCACACGGACTACATCAATGGCCAGATGATCTGGTATGACATGACCAGCCAAGAGAATTTTGATCGGCTTT
TTGCTACGGCATAGTTCTGACATTGTCTGGCTTTCTTTCCCTCCTTGGTCAATTTGGTGTGTATTCACTGA
TGGTCAGGAGCCTGATCAAGCCAGAGGAGAACCTCATGAGGACAGGCAACACAGCCCGAGCCAGGTCCATCC
GGACCATCCTACTGGTGTGTGGCCTCTTCACCTCTGTTTGTGCCCTTCCATATCACTCGCTCCTTCTACC
TCACCATCTGCTTCTGCTTCTCAGGACTGCCAGCTCTTGATGGCAGCCAGTGTGGCCTACAAGATATGG
AGGCCTCTGGTGAGTGTGAGCAGCTGCCTCAACCCAGTCTGTACTTCTTTCAAGGGGGGCAAAATAGAG
TCAGGCTCCTCCAGAACTGAGGCAGAACAAGTTGGGTGAGCATCCAGCTGGGAGGAAGAGATGCCAGGGT
TGAACAGATCTGGGTAAATGCCAAGGTGA

```

In a search of public sequence databases, the NOV43 nucleic acid sequence, located on chromosome 2, has 585 of 924 bases (63%) identical to a gb:GENBANK-ID:GDP2Y3|acc:X98283.1 mRNA from *Gallus gallus* (*G.domesticus* mRNA for G protein-coupled P2 receptor) ($E = 3.6e^{-45}$).

The disclosed NOV43 polypeptide (SEQ ID NO:166) encoded by SEQ ID NO:165 has 361 amino acid residues and is presented in Table 43B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV43 has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV43 may also localize to the mitochondrial inner membrane with a certainty of 0.5862, to

gi 13928944 ref NP_113868.1 (NM_031680)	purinergic receptor P2Y, G- protein coupled, 4; pyrimidinergic receptor P2Y, G- protein coupled, 4 [Rattus norvegicus]	361	118/310 (38%)	169/310 (54%)	6e-44
---	---	-----	------------------	------------------	-------

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 43D. In the ClustalW alignment of the NOV43 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 43D. ClustalW Analysis of NOV43

- 1) Novel NOV43 (SEQ ID NO:166)
- 2) gi|2829680|sp|P79928|P2Y8_XENLA P2Y PURINOCEPTOR 8 (P2Y8) (SEQ ID NO:517)
- 3) gi|2707256|gb|AAC60339.1| (AF031897) G protein coupled P2Y nucleotide receptor [Meleagris gallopavo] (SEQ ID NO:518)
- 4) gi|2495017|sp|Q98907|P2Y3_CHICK P2Y PURINOCEPTOR 3 (P2Y3) (NUCLEOSIDE DIPHOSPHATE RECEPTOR) (SEQ ID NO:519)
- 5) gi|10720180|sp|O93361|P2Y3_MELGA P2Y PURINOCEPTOR 3 (P2Y3) (NUCLEOSIDE DIPHOSPHATE RECEPTOR) (SEQ ID NO:520)
- 6) gi|13928944|ref|NP_113868.1| (NM_031680) purinergic receptor P2Y, G-protein coupled, 4; pyrimidinergic receptor P2Y, G-protein coupled, 4 [Rattus norvegicus] (SEQ ID NO:521)

			10	20	30	40	50	60	
25	NOV43	1	MLSILLPSRGSRSRSGRRG	ALLLEGASRDMKVD	MNTS	QEQGLC	QFSEKYKQVYLS	SLAYSI	60
	gi 2829680	1	-----MTEDIMATSYPT	LLTTPYLPKLLMNL	TNDTEDI	CVFDEG	CFKFLLPVSYSA		52
	gi 2707256	1	-----MDAPVRMFSLAP	WTPTPTP	--WLG	GNTTAAAEAK	CVFNEEFKFILL	ISYGI	50
	gi 2495017	1	-----MSMANETGGR	-----	-----	NSCTFH	EEFKQVLLPLVYSV		30
	gi 10720180	1	-----MSMANETAGR	-----	-----	NSCTFO	EEFKQVLLPLVYSV		30
30	gi 13928944	1	-----MTSAESLLFTS	-----	-----	LGPS	PSSG	GDGCRFNEEFKFLLPMSYAV	40
			70	80	90	100	110	120	
35	NOV43	61	IFILGLPLNGTVLWHSWG	QTKRWSCATT	YLVNLMVADLLYVL	LPFLITIT	YSLDDR	WPFPG	119
	gi 2829680	53	VFMVGLPLNIAAMWIFI	AKMRPWNPTTV	MFNLALSDTLYVL	SLPTLV	YYYADK	NNWPFPG	112
	gi 2707256	51	VFVVGGLPLNSWAMW	IFVSRMRPWNAT	TYMENLAISDTLYV	FSLPTLV	YYYADR	NNWPFPG	110
	gi 2495017	31	VFLGLPLNAVVGQI	WLARKALTRTT	TYMLNLA	MADLLYVCSL	PLLIYNY	TQKDYWPFPG	90
	gi 10720180	31	VFLGLPLNAVVGQI	WLARKALTRTT	TYMLNLA	TADLLYVCSL	PLLIYNY	TQKDYWPFPG	90
40	gi 13928944	41	VFVLGLALNAPTLL	WLFLRPLRP	WDATATYMFHL	ALSDTLYVL	SLPTLV	YYYAARNHWPFPG	100
			130	140	150	160	170	180	
45	NOV43	120	ELLCKLVHFLFYINLY	GSILLTLCISVH	QFLGVCHPLCS	IPYRT-RR	HANLGT	STTWALV	178
	gi 2829680	113	EVLCKLVRFIFYANLY	SSILFLTLCISV	HRYRGVCHP	ITSIRRMN	-AKHAY	ICALVWLSV	171
	gi 2707256	111	KVFCKLVRFIFYANLY	SSILFLTLCISV	HRYMGICH	PIRSIKWVK	-TKHAR	LHCVGVWLVV	169
	gi 2495017	91	DFTCKFVRFOFYTNL	HGSILFLTLCISV	ORYMGICH	PLASWHKKK	GKKLTW	LYCAAVWFLV	150
	gi 10720180	91	DFTCKFVRFOFYTNL	HGSILFLTLCISV	ORYMGICH	PLASWHKKK	GKKLTW	LYCAAVWFLV	150
	gi 13928944	101	TGLCKFVRFLFYWNLY	CSVLFLTLCISV	HRYLGICH	PLRATRWGR	-PRFAS	LCLGVWLVV	159

		190	200	210	220	230	240
5	NOV43	179	VLQLLEPLAFSTHSDYINGQMIWYDMISQENFDRLFAVGIVLTISGF	LSLLGHFGVYSLM	237		
	gi 2829680	172	TLCLVKNLIFFVTVSPKVNKNTICHDTTRPEDFARYVEYSTALMCLLFGTFCPLITAGCYGLM	231			
	gi 2707256	170	TICLIENLIFFVITSSKDNSTLCHDTIKPEEFDHYVHYSSSIMALLFGIPFLVITVVCYCLM	229			
	gi 2495017	151	IAQCLPTFFVFASTGTQRNRTVCYDLSPPDRSTSYFFPGITLTITITGFLPFAAILIACYCSM	210			
	gi 10720180	151	IAQCLPTFFVFASTGTQRNRTVCYDLSPPDRSASYFFPGITLTITITGFLPFAAILIACYCSM	210			
	gi 13928944	160	AGCLVENLFFVITNANGTTILCHDTILPEEFDHYVYESSAVMVLLECLPFLITITVVCYGLM	219			
10		250	260	270	280	290	300
	NOV43	238	VRSLIKPEENLMRTG-NTARARSIRTHLLVCGLETLICFVFPFHITRSEYLTICFLLSQDCQ	296			
	gi 2829680	232	TRELMKPIVSGNQOTLPSYKKRSIKTHFVMIARFAICFMPFHITRILYYAR-LLGIKCY	290			
	gi 2707256	230	AKRLCKRSFPSPSPRPVPSYKKRSIKMIIIVLTVEAICFVFPFHITRILYYTSR-YFOADCQ	288			
15	gi 2495017	211	ARILCQKDELIGLA-VHKKKDKAVRMIIIVIVFISISFEPFHILTKTIYLLIVRSSASLPCP	269			
	gi 10720180	211	ARILCQKDELIGLA-VHKKKDKAVRMIIIVIVFISISFEPFHILTKTIYLLIVRSSPTLPCP	269			
	gi 13928944	220	ARRLYRPLPGAGQ---SSRLRSRLTIAVVLTVFAVCFVFPFHITRITIIYQAR-LLQADCH	275			
20		310	320	330	340	350	360
	NOV43	297	LLMAAQCGQLQDMEASGECEQLPQPSVLSFKGCKNRVRLQKLFQ	---	---	---	342
	gi 2829680	291	ALNVINVTYKVTREPLASANSCLDP-ILYFLANDRYRRRLIRTVRRRSSVPNRRMCHTNHP	---	---	---	349
	gi 2707256	289	TLNIINFTYKIIITRPLASINSCLDP-ILYFMAGDKYRGRLRRGAAQR----	P	---	---	334
25	gi 2495017	270	TLQAFATAYKCTRPFASMSVLDLP-ILFYFTQRKFR-ESTR	---	---	---	308
	gi 10720180	270	TLQAFATAYKCTRPFASMSVLDLP-ILFYFTQRKFR-ESTR	---	---	---	308
	gi 13928944	276	VLNIVNVVYKVTREPLASANSCLDP-VLYLFTGDKYRNQLQQLCRGS-	---	---	---	320
30		370	380	390	400	410	420
	NOV43	342	---	KLGEHP	---	AGRKRCPGILNR	359
	gi 2829680	350	QTEPHMTAGPLPVISAAEIPSNGSMVRDENEGSGREHRVEWTDITKEINQMMNRRSTIKRN	---	---	---	409
	gi 2707256	335	RPVP-SSLALVS---	PSVDSSVVGSCCN-	SES	RGMTVWS	370
	gi 2495017	308	---Y---LLDKMS	---	SKWRQDHCISY	---	326
	gi 10720180	308	---Y---LLDKMS	---	SKWRHDHCITY	---	326
35	gi 13928944	321	KPKPRTAASSLALVT---	LHEESISR-	---	WADTHQDSTFSA	355
40		430	440	450	460	470	480
	NOV43	359	---	SG	---	---	361
	gi 2829680	410	STDKNDMKENRHGENYLPYVEVVEKEDYETKRENKKTTEQSSKTNAEQDELQTQIDSRLK	---	---	---	469
	gi 2707256	370	---	RCGQ	---	---	374
	gi 2495017	326	---	GS	---	---	328
	gi 10720180	326	---	GS	---	---	328
45	gi 13928944	355	---	YECDRL	---	---	361
50		490	500	510	520	530	540
	NOV43	361	---	---	---	---	361
	gi 2829680	470	RGKWQLSSKKGAQAQENEKGHMEPSFEGEGTSTWNLLTPKMYGKKDR	---	---	---	529
	gi 2707256	374	---	---	---	---	374
	gi 2495017	328	---	---	---	---	328
	gi 10720180	328	---	---	---	---	328
	gi 13928944	361	---	---	---	---	361
55		361	---	---	---	---	
	NOV43	361	---	---	---	---	
	gi 2829680	530	ELQNFPKA	---	---	---	537
	gi 2707256	374	---	---	---	---	374
60	gi 2495017	328	---	---	---	---	328
	gi 10720180	328	---	---	---	---	328
	gi 13928944	361	---	---	---	---	361

65 Tables 43E lists the domain descriptions from DOMAIN analysis results against NOV43. This indicates that the NOV43 sequence has properties similar to those of other proteins known to contain this domain.

Table 43E Domain Analysis of NOV43

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 85.0% aligned
Score = 111 bits (277), Expect = 8e-26

	NOV43:	69	NGTVLWHSWGQTKRWSCATTYLVNLMVADLLYVL-LPFLIITYSLDDRWPFGELLCKLVH	127
5	Sbjct:	2	NLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPWALYYLVGGDWVFGDALCKLVG	61
	NOV43:	128	FLFYINLYGSILLTLCISVHQFLGVCHPLCSLPYRTRRHAWLGTSTTVALVVLQ-LLPTL	186
10	Sbjct:	62	ALFVVNGYASILLTLCISVHQFLGVCHPLCSLPYRTRRHAWLGTSTTVALVVLQ-LLPTL	121
	NOV43:	187	AFSHTDYINGQMIWYDMSQENFDRFLFAYGIVLTLSGFLSLLGHFVYSLMVRSLIKPE-	245
	Sbjct:	122	FSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPPLLVLVCYTRILRTLKRAR	181
15	NOV43:	246	-ENLMRTGNTARARSIRTILLVCGLFTLCFVPFHIT	280
	Sbjct:	182	SQSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIV	217

The P2Y Purinoreceptor belongs to the family of G-Protein Coupled Receptors. G-protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide range of functions (including various autocrine, paracrine and endocrine processes). They show considerable diversity at the sequence level, on the basis of which they can be separated into distinct groups. We use the term clan to describe the GPCRs, as they embrace a group of families for which there are indications of evolutionary relationship, but between which there is no statistically significant similarity in sequence [1]. The currently known clan members include the rhodopsin-like GPCRs, the secretin-like GPCRs, the cAMP receptors, the fungal mating pheromone receptors, and the metabotropic glutamate receptor family. The rhodopsin-like GPCRs themselves represent a widespread protein family that includes hormone, neurotransmitter and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins. Although their activating ligands vary widely in structure and character, the amino acid sequences of the receptors are very similar and are believed to adopt a common structural framework comprising 7 transmembrane (TM) helices. See CMKRL2 (601805). Using degenerate PCR to find cDNAs encoding new G protein coupled-receptors in human B cells, Owman et al. (1996) identified a CMKRL1 cDNA which encodes a 352-amino acid polypeptide with a calculated mass of 43 kD. The nearest homologs of this novel sequence are the chemoattractant leukocyte receptors, such as the C5a anaphylatoxin receptor and the FMLP receptor. Northern blotting revealed transcripts of 5 kb and 7.5 kb in several tissues of the immune system including spleen, thymus, and lymph node. Owman et al. (1996) considered the high level of expression in

lymphoid tissues suggestive of the role of CMKRL1 in the regulation of the inflammatory system. The authors mapped the CMKRL1 gene to 14q11.2-q12 by fluorescence in situ hybridization. Akbar et al. (1996) used a chicken P2Y3 cDNA to screen a human erythroleukemia (HEL) cell cDNA library and cloned a purinoceptor cDNA, which they
5 termed P2Y7. Sequencing revealed an open reading frame coding for a polypeptide of 352 amino acids having 7 putative transmembrane domains. The P2Y7 receptor has 23 to 30% identity to other P2Y receptors, but forms a unique branch within the P2Y family. Northern blot analysis showed that the P2Y7 gene produced a 1.6-kb transcript which is expressed at highest levels in human heart, human skeletal muscle, rat heart, and rat cardiomyocytes and at
10 lower levels in human brain and human liver. Akbar et al. (1996) noted that its expression in HEL cells is below the threshold of detection by Northern blot. Binding and displacement assays in COS-7 cells showed that P2Y7 has a high affinity for ATP and much less for UTP and ADP. The rank order of affinities in the binding series was distinct from any known for the P2Y1-P2Y6 receptors. Like other P2Y receptors, P2YR is coupled to phospholipase C and
15 not to adenylate cyclase. Akbar et al. (1996) speculated that P2Y7 may be the cardiac P2Y receptor involved in the regulation of cardiac muscle contraction through modulation of L-type calcium currents. Akbar et al. (1996) used PCR on a panel of mouse-rodent somatic cell hybrids to localize the P2RY7 gene to human chromosome 14. Somers et al. (1997) did sequence tagged site (STS) mapping of the P2RY7 gene using the National Center for
20 Biotechnology Information (NCBI) database. In this way, they positioned the P2RY7 gene between D14S283 and D14S264. Leukotriene B4 (LTB4) is a potent chemoattractant that is primarily involved in inflammation, immune responses, and host defense against infection (Samuelsson et al., 1987; Chen et al., 1994). LTB4 activates inflammatory cells by binding to its cell surface receptor, BLTR. LTB4 can also bind and activate the intranuclear transcription
25 factor PPAR-alpha, resulting in the activation of genes that terminate inflammatory processes (Devchand et al., 1996). Yokomizo et al. (1997) cloned the cDNA encoding a cell surface LTB4 receptor that is highly expressed in human leukocytes. Two cDNA clones isolated from retinoic acid-differentiated HL-60 cells contained identical open reading frames encoding a protein of 352 amino acids and predicted to contain 7 membrane-spanning domains, but
30 different 5-prime untranslated regions. In Chinese hamster ovary (CHO) cells stably expressing this receptor, LTB4 induced increases in intracellular calcium, accumulation of D-myo-inositol-1,4,5-triphosphate, and inhibition of adenylyl cyclase. Furthermore, CHO cells expressing exogenous BLTR showed marked chemotactic responses toward low concentrations of LTB4 in a pertussis-toxin-sensitive manner. Yokomizo et al. (1997) found

that the putative purinoceptor P2Y7 has a primary structure identical to that of one of the BLTR clones, HL-5. To determine whether BLTR also functions as a purinoceptor, they established stable transformants of BLTR in glioma cells that possess negligible amounts of intrinsic purinoceptors. In these cells, up to 300 microM caused no change in intracellular calcium levels, but significant increases in the calcium concentrations were induced by exposure to 10 nanoM LTB4. These results were interpreted to indicate that this receptor is not a purinoceptor, but a BLTR.

The disclosed NOV43 nucleic acid of the invention encoding a P2Y Purinoceptor -like protein includes the nucleic acid whose sequence is provided in Table 43A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 43A while still encoding a protein that maintains its UDP[Glycosyltransferase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV43 protein of the invention includes the P2Y Purinoceptor -like protein whose sequence is provided in Table 43B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 43B while still encoding a protein that maintains its P2Y Purinoceptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 62 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this P2Y Purinoceptor -like protein (NOV43) is a member of a "P2Y Purinoceptor family". Therefore, the NOV43 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential

therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV43 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, and/or other diseases and pathologies.

NOV43 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV43 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below. The disclosed NOV43 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV44

A disclosed NOV44 nucleic acid of 934 nucleotides (also referred to as CG56692-01) encoding a G Protein Coupled Receptor-like protein is shown in Table 44A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 15-17 and ending with a TAA codon at nucleotides 921-923. The start and stop codons are shown in bold in Table 44A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 44A. NOV44 nucleotide sequence (SEQ ID NO:167).

TGACCTTGGGAATCT**AT**GGACATACCAAAAATATCACAGAATTTTTTCATGCTGGGGCTCTCACAGAACTCAG
 AGGTACAGAGAGTTCTCTTTGTGGTCTTTTTGCTGATCTATGTGGTCACGGTTTGTGGCAACATGCTCATTG
 TGGTCACTATCACCTCCAGCCCCACGCTGCGCTTCCCCGTGTGTAATTTTTCTGGCCAACTCATCTTTATCT
 ACACCTTTTATTCTTCTTATGCTTCTTAAGCTTCT**AA**CTCATTGGTACTCATTGTATGAGGGGAGAACCCTATT
 ATGAGTGCTGTCATGGCTCAGCTC**TT**TGGAGCTCAATTTTTGGAGGAGGTGTTGATATCATTCTGCTCACAGTGA
 TGGCTTATGACCGCTATGTGGCCATCTGTAAGCCCTGCACAATACTACCATCATGACCAGGCATCTCTGTG
 CCATGCTTGTAGGGGTGGCTTGGCTTGGGGGCTTCTCGTCACTATTGGTTACGCTCCCTCGTGGTCTTTGGT
 TGCCTTCTGTGGGCCCAATGTGATCAATCACTTTGGCTGTGACTTGTAACCTTTGCTGGAAGTTGCCTGCA
 CCAATACGTATGTCATTGGTCTGCTGGTGGTTGCCAACAGTGGTTAATCTGCCTGTTGAACCTTCCTCATGC
 TGGCTGCCTCCTACATTGTCATCCTGTACTCCTTGAGGTTCCACAGTGCAGATGGGAGATGCAAAGCCCTCT
 CCACCTGTGGAGCCCACTTCAATGTTGTTGGCCTGTTCTTTTGCCCTGTATATTTACTATGTGATCATCAT
 TTTCTACTGTTTACCTATAGACAAAGATATGGCATTATTTTATGGTATTTCTGACACCTATGTGTAATCCCATCA
 TTTTATACCTTGAGAAATGAAGAGGTTAAAAAATGCCATGAGAAAGCTCTTTACATGGT**TA**GAAATTGCAGG

In a search of public sequence databases, the NOV44 nucleic acid sequence, located on chromosome 7, has 783 of 920 bases (85%) identical to a gb:GENBANK-ID:AB030895|acc:AB030895.1 mRNA from *Mus musculus* (gene for odorant receptor MOR18, complete cds) ($E = 4.5e^{-146}$).

The disclosed NOV44 polypeptide (SEQ ID NO:168) encoded by SEQ ID NO:167 has 302 amino acid residues and is presented in Table 44B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV44 has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV44 may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the mitochondrial inner membrane with a certainty of 0.0300. The most likely cleavage site for NOV44 is between positions 39 and 40: VCG-NM.

Table 44B. Encoded NOV44 protein sequence (SEQ ID NO:168).

MDIPQNITEFFMLGLSQNSEVQRVLFVVFLLIYVVTVCGNMLIVVTITSSPTLASPVYFFLANLSFIDTFYS
SSMAPKLIADSLYEGRTISYECCMAQLFGAHFLGGVEIILLTVMAYDRYVAICKPLHNTTIMTRHLCAMLVG
VAWLGGFLHSLVQLLLVLWLPFCGPNVINHFACDLYPLLEVACTNTYVIGLLVVANSGLICLLNFLMLAASY
IVILYSLRSHSADGRCKALSTCGAHFIVVALFFVPCIFTYVHPFSTLPIDKNMALFYGILTPMLNPLIYTLR
NEEVKNAMRKLFTH

A search of sequence databases reveals that the NOV44 amino acid sequence has 257 of 301 amino acid residues (85%) identical to, and 280 of 301 amino acid residues (93%) similar to, the 308 amino acid residue ptmr:SPTREMBL-ACC:Q9R0K2 protein from *Mus musculus* (Mouse) (Odorant Receptor MOR18) ($E = 5.0e^{-138}$).

NOV44 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 44C.

Table 44C. BLAST results for NOV44

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17472367 ref XP_061659.1 (XM_061659)	similar to odorant receptor 16 (H. sapiens) [Homo sapiens]	324	302/302 (100%)	302/302 (100%)	e-143
gi 11496249 ref NP_067343.1 (NM_021368)	odorant receptor 16 [Mus musculus]	308	257/301 (85%)	280/301 (92%)	e-127
gi 11464995 ref NP_065261.1 (NM_020515)	gene for odorant receptor A16 [Mus musculus]	302	234/300 (78%)	262/300 (87%)	e-111
gi 17459946 ref XP_062088.1 (XM_062088)	similar to odorant receptor 16 (H. sapiens) [Homo sapiens]	316	191/295 (64%)	232/295 (77%)	9e-95

gi 17472365 ref XP_061658.1 (XM_061658)	similar to odorant receptor 16 (H. sapiens) [Homo sapiens]	544	183/296 (61%)	230/296 (76%)	1e-91
--	--	-----	---------------	---------------	-------

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 44D. In the ClustalW alignment of the NOV44 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 44D. ClustalW Analysis of NOV44

- 1) Novel NOV44 (SEQ ID NO:168)
- 2) gi|17472367|ref|XP_061659.1| (XM_061659) similar to odorant receptor 16 (H. sapiens) [Homo sapiens] (SEQ ID NO:522)
- 3) gi|11496249|ref|NP_067343.1| (NM_021368) odorant receptor 16 [Mus musculus] (SEQ ID NO:523)
- 4) gi|11464995|ref|NP_065261.1| (NM_020515) gene for odorant receptor A16 [Mus musculus] (SEQ ID NO:524)
- 5) gi|17459946|ref|XP_062088.1| (XM_062088) similar to odorant receptor 16 (H. sapiens) [Homo sapiens] (SEQ ID NO:525)
- 6) gi|17472365|ref|XP_061658.1| (XM_061658) similar to odorant receptor 16 (H. sapiens) [Homo sapiens] (SEQ ID NO:526)

			10	20	30	40	50	60	
25	NOV44	1						1
	gi 17472367	1	----- ----- ----- ----- ----- ----- ----- ----- -----						1
	gi 11496249	1	----- ----- ----- ----- ----- ----- ----- ----- -----						1
	gi 11464995	1	----- ----- ----- ----- ----- ----- ----- ----- -----						1
	gi 17459946	1	----- ----- ----- ----- ----- ----- ----- ----- -----						1
30	gi 17472365	1	MTQISSNAFSRDFQNSNAFEVQVKDPIHVEDVPGPKSEFCVFPSTPQASGNFQNFQD	60					
			70	80	90	100	110	120	
35	NOV44	1	----- ----- ----- ----- ----- ----- ----- ----- -----						
	gi 17472367	1	MHEWLFRFKGNTAPAFSVTLESMDIPONITEFFMLGLSQNSEVQRVLFVVFLLIYVVTVC	38					60
	gi 11496249	1	----- ----- ----- ----- ----- ----- ----- ----- -----						38
	gi 11464995	1	----- ----- ----- ----- ----- ----- ----- ----- -----						38
	gi 17459946	1	----- ----- ----- ----- ----- ----- ----- ----- -----						38
	gi 17472365	61	STEAIPLDEDQRKINYPNTKLDFOVMNITEFILLGLTONAEAKLLFAVETLIYFLTMV	120					
40			130	140	150	160	170	180	
45	NOV44	39	GNMLIVVTITSSPTLASPMYFFFLANLSFIDTFYSSSMAPKLIADSLYEGRTISYECCMA	97					
	gi 17472367	61	GNMLIVVTITSSPTLASPMYFFFLANLSFIDTFYSSSMAPKLIADSLYEGRTISYECCMA	119					
	gi 11496249	39	GNMLIVVTITSSPTLASPMYFFFLANLSFIDTFYSSSMAPKLIADSLYEGRTISYECCMA	97					
	gi 11464995	39	GNMLIVVTITSSPTLASPMYFFFLANLSFIDTFYSSSMAPKLIADSLYEGRTISYECCMA	97					
	gi 17459946	39	GNMLIVVTITSSPTLASPMYFFFLANLSFIDTFYSSSMAPKLIADSLYEGRTISYECCMA	98					
	gi 17472365	121	DNELIVVTITITSPA-LDSPVYFFFLSFFSFIDGCSSTMAPKMIFFDLLTEKKTISFSGCMT	179					
50			190	200	210	220	230	240	
	NOV44	98	QLFGAHLGGVEIILLTVMAYDRYVAICKPLHNTTIMTRHLCAMLVGVAVLGGFLHSLVC	157					
	gi 17472367	120	QLFGAHLGGVEIILLTVMAYDRYVAICKPLHNTTIMTRHLCAMLVGVAVLGGFLHSLVC	179					


```
gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin
family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 82.0 bits (201), Expect = 4e-17
```

5

10

15

20

25

35

40

functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures
5 include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic
10 acids, and their complements, up to about 15 percent of the bases may be so changed.

The disclosed NOV44 protein of the invention includes the G Protein Coupled Receptor -like protein whose sequence is provided in Table 44B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 44B while still encoding a protein that maintains its G Protein Coupled
15 Receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 39 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this G Protein Coupled Receptor -like
20 protein (NOV44) is a member of a "G Protein Coupled Receptor family". Therefore, the NOV44 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug
25 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV44 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Systemic lupus erythematosus, Autoimmune disease,
30 Asthma, Emphysema, Scleroderma, allergy, ARDS, and/or other diseases and pathologies.

NOV44 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV44 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-

NOVX Antibodies" section below. The disclosed NOV44 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV45

A disclosed NOV45 nucleic acid of 994 nucleotides (also referred to as CG56694-01) encoding a Mas Proto-Oncogene-like protein is shown in Table 45A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 17-19 and ending with a TGA codon at nucleotides 980-982. The start and stop codons are shown in bold in Table 45A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 45A. NOV45 nucleotide sequence (SEQ ID NO:169).

```

ACTAGGGTTCCTGAGCATGGATCCAACCATCCAGCCTTGGGTACAGAACTGACACCAATCAATGGACGGGA
GGAGACTCCTTGCTACAAGCAAACCTGAGCCTCACAGGGCTGACGTGCATCGTTCCCTTGTCGGGATGAC
AGGAAATGCAGTCGTGCTCTGGCTCCTGGGCTTCCGCATGCGCAGGAACGCCTTCTCCATCTACATCTTCAA
CCTGTCCATGGCCGACTTCCTCTTTCTCAGAAGCCACATTATACGTTTCCGTTAAGCCTCATCAATATCCT
CCTATCCCATCTTCAAAATCCTCAGCCCTGTGATGATGTTTTCTACCTTGCAAGCCTGAGCTTTCTAAGCGC
CATGAGCACCGAGCGCTGCCTGTACGTCCTGTGGCCCATCTGGGAGCGCTGCCGCCCCCGCCCTACACCTG
TCAGCGGTTCGTGTGTCATGCTCTGGGCCCTGTCTCTGCTGCGGAGCGTCCCTGGAGTGGAGTTTCTGTGAC
TTCTGTTTTAGTGGTGCTGATTCTGTTTGGTGTAACATCAGATTTCATCATAGTAGGGGGGCTGATTTT
TTATGTGTGGCTCTCTGTGGTTCCAGCCTGGTCTGCTGGTCAGGATCCTTTGTGGGTCCCGGAAGATGCCA
CTGACCAGGCTGTACGTGACCATCCTGCTCATAGCGCTGGTCTTCTCTCTGTGGCCTGCCCTTTGGCATT
CGGTTTTTCTATTTTCATGGAACACGTGGATTGGAAGTCTTATATTGTACGTTTCATCTAGTTTCATT
TTCTTTCTCTTTAACGGCCAACCCCAACATTTACTTCTTCGTGGGCTCCTTAAGGAGTGTCAAAAAAG
GCAGAACCTGAAGCTGGTTCTCCAGAGGGCTCTGCAGGACACGACTGAGGTGAATGAAGTGGACGATGGCT
TCCTGAGGAAACCTGGAGCTGTGAGGAAGCAGATTGGGCAGTAGGAAAGAGCCTCT

```

In a search of public sequence databases, the NOV45 nucleic acid sequence, located on chromosome 7, has 353 of 580 bases (60%) identical to a gb:GENBANK-
ID:108606|acc:108606.1 mRNA from Unknown. (Sequence 1 from Patent WO 8707472) ($E = 1.2e^{-10}$).

The disclosed NOV45 polypeptide (SEQ ID NO:170) encoded by SEQ ID NO:169 has 319 amino acid residues and is presented in Table 45B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV45 has no signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6400. Alternatively, NOV45 may also localize to the Golgi body with a certainty of 0.4600, to the endoplasmic reticulum (membrane) with a certainty of 0.3700, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000.

Table 45B. Encoded NOV45 protein sequence (SEQ ID NO:170).

```

MIQLLITSGGLCSQGHQTRVPEHGSNHPSLGYRTDNTQWTGGDSLLQANPEPHRADVHRFPDRDRKCSRALA

```

PGLPHAQERLLHLHLQPVHGRPLSLQKPHYTFSVKPHQYPPSHLQNPQPCDDVFLPCKPELSKRHEHRALPV
 RPAHLGALPPPPLHLSAVVCVMLWALSLLRSVLEWSFCDFLFSGADSVWCKTSDFIIVGGLIFLCVALCGS
 SLVLLVRILCGSRKMPLTRLYVTILLIALVFLLCGLPFGIRFFLFSWNHVDLEVLYCHVHLVSIPLSSLNGQ
 PQHLLLRGLLKAVSKKAEPEAGSPEGSAGHD

A search of sequence databases reveals that the NOV45 amino acid sequence has 50 of 168 amino acid residues (29%) identical to, and 87 of 168 amino acid residues (51%) similar to, the 378 amino acid residue ptnr:SWISSPROT-ACC:P35410 protein from *Homo sapiens* (Human) (Mas-Related G Protein-Coupled Receptor MRG) ($E = 5.0e-138$).

NOV45 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 45C.

Table 45C. BLAST results for NOV45					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15546062 gb AAK91804.1 (AY042213)	MrgX1 G protein-coupled receptor [<i>Homo sapiens</i>]	322	108/130 (83%)	116/130 (89%)	2e-56
gi 17472340 ref XP_061650.1 (XM_061650)	similar to MrgX1 G protein-coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	1589	108/130 (83%)	116/130 (89%)	3e-56
gi 16876453 ref NP_473372.1 (NM_054031)	G protein-coupled receptor MRGX3 [<i>Homo sapiens</i>]	322	104/130 (80%)	118/130 (90%)	2e-53
gi 17461239 ref XP_062249.1 (XM_062249)	similar to MrgX3 G protein-coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	322	103/130 (79%)	118/130 (90%)	7e-53
gi 16876455 ref NP_473373.1 (NM_054032)	G protein-coupled receptor MRGX4 [<i>Homo sapiens</i>]	322	97/124 (78%)	104/124 (83%)	2e-46

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 45D. In the ClustalW alignment of the NOV45 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 45D. ClustalW Analysis of NOV45

- 1) Novel NOV45 (SEQ ID NO:170)
- 2) gi|15546062|gb|AAK91804.1| (AY042213) MrgX1 G protein-coupled receptor [*Homo sapiens*] (SEQ ID NO:527)

5

15

25

35

45

55

65

390

[illegible]

[illegible]

	NOV45	319		319
	gi 15546062	322		322
	gi 17472340	1321	ISFLSVYVEISPKLHNTKGSSEVTESQKGSTSLARGSTSSSTLDRRRRKDAQQSHIEPHFK	1380
	gi 16876453	322		322
	gi 17461239	322		322
	gi 16876455	322		322
			1390 1400 1410 1420 1430 1440	
			
10	NOV45	319		319
	gi 15546062	322		322
	gi 17472340	1381	GTLVVNLRGTRLGFLSMNPTIPALDTEIAPISDTEETHPHRCGMEVLVLIVLILIIDLVG	1440
	gi 16876453	322		322
	gi 17461239	322		322
15	gi 16876455	322		322
			1450 1460 1470 1480 1490 1500	
			
20	NOV45	319		319
	gi 15546062	322		322
	gi 17472340	1441	LAGNAVMLWLLGFCMHSNTFSLYILNLARADFLCTCFQIITFINFFSDFVSSLSIHFSRF	1500
	gi 16876453	322		322
	gi 17461239	322		322
25	gi 16876455	322		322
			1510 1520 1530 1540 1550 1560	
			
30	NOV45	319		319
	gi 15546062	322		322
	gi 17472340	1501	VTTVLFSACITGLSMLSTISTEHRSLVLPICSANPIIYFFMGSFRLQNRKTLKLVLR	1560
	gi 16876453	322		322
	gi 17461239	322		322
	gi 16876455	322		322
			1570 1580	
			
35	NOV45	319		319
	gi 15546062	322		322
	gi 17472340	1561	ALQDMLEVDEGGGQLPEETLKLSGSRLGP	1589
	gi 16876453	322		322
	gi 17461239	322		322
40	gi 16876455	322		322

Table 45E lists the domain description from DOMAIN analysis results against NOV45. This indicates that the NOV45 sequence has properties similar to those of other proteins known to contain this domain.

Table 45E Domain Analysis of NOV45	
gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)	
CD-Length = 254 residues, only 55.9% aligned	
Score = 41.2 bits (95), Expect = 9e-05	

NOV44:	165	CVMLWALSLLRSVLE-----WSFCDFLFGADSVWCKTSDFIIVGGLIFLCV	211
		+++ + + + + + + +	
Sbjct:	105	ILLVWVLALLLSLPPLLFSLWLRRTVEEGNTTVCLIDFPESVKRSYVLLSTLVGFVLPLL	164

NOV44: 212 ALCGSSLVLL---VRILCGSRKMPL---TRLYVTILLIALVFLLCGLPFGIRFFLFSWN 264
 | + + | | + +|++ +||+|| + | |
 Sbjct: 165 ILVCYTRILRTLKRKARSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSLC 224

5 NOV44: 265 HVDLEVLYCHVHLVSIFLSSLN 286
 + + + |++++|+ +|
 Sbjct: 225 LLSIWRVLPTALLITLWLAYVN 246

10 The Mas Proto-Oncogene belongs to the family of G-Protein Coupled Receptors. G-protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide range of functions (including various autocrine, paracrine and endocrine processes). They show considerable diversity at the sequence level, on the basis of which they can be separated into distinct groups. We use the term clan to describe the GPCRs, as they embrace a group of

15 families for which there are indications of evolutionary relationship, but between which there is no statistically significant similarity in sequence [1]. The currently known clan members include the rhodopsin-like GPCRs, the secretin-like GPCRs, the cAMP receptors, the fungal mating pheromone receptors, and the metabotropic glutamate receptor family.

The human mas oncogene was originally detected by its ability to transform NIH 3T3

20 cells. We previously showed that the protein encoded by this gene is unique among cellular oncogene products in that it has seven hydrophobic potential transmembrane domains and shares strong sequence similarity with a family of hormone-receptor proteins. We have now cloned the rat homolog of the mas oncogene, determined its DNA sequence, and examined its expression in various rat tissues. A comparison of the predicted sequences of the rat and

25 human mas proteins shows that they are highly conserved, except in their hydrophilic amino-terminal domains. Our examination of the expression of mas, determined by RNA-protection studies, indicates that high levels of mas RNA transcripts are present in the hippocampus and cerebral cortex of the brain, but not in other neural regions or in other tissues. This pattern of expression and the similarity of mas protein to known receptor proteins suggest that mas

30 encodes a receptor that is involved in the normal neurophysiology and/or development of specific neural tissues.

The rhodopsin-like GPCRs themselves represent a widespread protein family that includes hormone, neurotransmitter and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins. Although their

35 activating ligands vary widely in structure and character, the amino acid sequences of the receptors are very similar and are believed to adopt a common structural framework comprising 7 transmembrane (TM) helices

The disclosed NOV45 nucleic acid of the invention encoding a Mas Proto-Oncogene-like protein includes the nucleic acid whose sequence is provided in Table 45A or a fragment

thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 45A while still encoding a protein that maintains its Mas Proto-Oncogene -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 40 percent of the bases may be so changed.

The disclosed NOV45 protein of the invention includes the Mas Proto-Oncogene-like protein whose sequence is provided in Table 45B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 45B while still encoding a protein that maintains its Mas Proto-Oncogene-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Mas Proto-Oncogene -like protein (NOV45) is a member of a "Mas Proto-Oncogene family". Therefore, the NOV45 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV45 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-

telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, and/or other diseases and pathologies.

NOV45 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV45 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV45 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV46

NOV46 includes three novel Mas Proto-Oncogene -like proteins disclosed below. The disclosed sequences have been named NOV46a and NOV46b.

NOV46a

A disclosed NOV46a nucleic acid of 997 nucleotides (also referred to as CG56696-01) encoding a Mas Proto-Oncogene-like protein is shown in Table 46A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 12-14 and ending with a TGA codon at nucleotides 978-980. The start and stop codons are shown in bold in Table 46A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 46A. NOV46a nucleotide sequence (SEQ ID NO:171).

```
GGTTTCTGAGCATGGATCCAACCATCTCAACCTTGGACACAGAACTGACACCAATCAACGGAAGTGGAGAGA
CTCTTTGCTACAAGCAGACCTTGAGCCTCACGGTGCTGACGTGCATCGTTCCCTTGTCGGGCTGACAGGAA
ACGCAGTTGTGCTCTGGCTCCTGGGCTGCCGATGCGCAGGAACGCCCTTCTCCATCTACATCTCAACTTGG
CCGCAGCAGACTTCCTCTCTCTCAGCGGCCGCCTATATATTCCTGTTAAGCTTCATCAGTATCCCCCATA
CCATCTCTAAAATCCTCTATCCTGTGATGATGTTTTCTACTTTGCAGGCCTGAGCTTCTGAGTGCCGTGA
GCACCGAGCGCTGCCTGTCCGTCTGTGGCCCATCTGGTACCGCTGCCACCGCCCCACACACCTGTGACGCG
TGGTGTGTGTCCTGCTCTGGGCCCTGTCCCTGCTGCGGAGCATCCTGGAGTGGATGTTATGTGGCTTCCTGT
TCAGTGGTGCTGATTCTGCTTGGTGTCAAACATCAGATTTTCATCACAGTCGCGTGGCTGATTTTTTATGTG
TGGTTCTCTGTGGGTCCAGCCTGGTCTGCTGATCAGGATTCTCTGTGGATCCCGGAAGATACCGCTGACCA
GGCTGTACGTGACCATCTGCTCACAGTACTGGTCTTCTCCTCTCTGTGGCCTGCCCTTTGGCATTGAGTTTT
TCCTATTTTTATGGATCCAGTGGACAGGGAAGTCTTATTTGTGTCATGTTTCATCTAGTTTCTATTTTCCTGT
CCGCTCTTAACAGCAGTGCCAACCCCATCATTTACTTCTTCGTGGGCTCCTTTAGGCAGCGTCAAAATAGGC
AGAACCTGAAGCTGGTTCTCCAGAGGGCTCTGCAGGACGCGTCTGAGGTGGATGAAGGTGGAGGCGAGCTTC
CTGAGGAAATCCTGGAGCTGTGCGGAAGCAGATTGGAGCAGTGAAGGAAGAGCCTCTGCCCT
```

In a search of public sequence databases, the NOV46a nucleic acid sequence, located on chromosome 7, has 430 of 705 bases (60%) identical to a gb:GENBANK-ID:MMU249895|acc:AJ249895.1 mRNA from *Mus musculus* (mas proto-oncogene and Igf2r gene for insulin-like growth factor type 2 and L41ps and Au76 pseudogenes) ($E = 9.3e^{-22}$).

The disclosed NOV46a polypeptide (SEQ ID NO:172) encoded by SEQ ID NO:171 has 322 amino acid residues and is presented in Table 46B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV46a has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV46a may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV46a is between positions 45 and 46: TGN-AV.

Table 46B. Encoded NOV46a protein sequence (SEQ ID NO:172).	
MDPTISTLDTLTPINGTEETLCYKQTLSTLVTLCIVSLVGLTGNAVVLWLLGCRMRRNAFSIYILNLAAAD	
FLFLSGRLIYSLLSFISIPHTISKILYPVMMFSYFAGLSFLSAVSTERCLSVLWPIWYRCHRPHTLSAVVCV	
LLWALSLLRSILEWMLCGFLFSGADSAWCQTSDFITVAWLIFLCVVLGSSSLVLLIRILCGSRKIPLTRLIV	
TILLTVLVFLLCGLPFGIOFFFLWIHVDREVLFCVHLVSIFLSALNSSANPIIYFFVGSFRQRQNRQNLK	
LVLQALQDASEVDEGGQLPEEILELSGSRLEQ	

A search of sequence databases reveals that the NOV46a amino acid sequence has 110 of 275 amino acid residues (40%) identical to, and 175 of 275 amino acid residues (63%) similar to, the 324 amino acid residue ptnr:SWISSPROT-ACC:P12526 protein from *Rattus norvegicus* (Rat) (Mas Proto-Oncogene) ($E = 1.6e^{-45}$).

NOV46b

In the present invention, the target sequence identified previously, NOV46a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high

redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV46b. This differs from the previously identified sequence (NOV46a) in having 2 amino acid changes.

A disclosed NOV46b nucleic acid of 964 nucleotides (also referred to as CG56696-02) encoding a Mas-Related G Protein-Coupled Receptor -like protein is shown in Table 46C. An open reading frame was identified beginning with a ACC initiation codon at nucleotides 3-5 and ending with a TGA codon at nucleotides 960-962. The start and stop codons are shown in bold in Table 46C, and the 5' and 3' untranslated regions, if any, are underlined. Because the start codon is not a traditional initiation codon, NOV46b could be a partial reading frame extending further in the 5' direction.

Table 46C. NOV46b nucleotide sequence (SEQ ID NO:173).

CAACCATCTCAACCTTGGACACAGAACTGACACCAATCAACGGAAGTGGAGAGACTCTTTGCTACAAGCAGACCTTGAGCCTCACGGTGTGACGTGCATCGTTCCCTTGTGCGGCTGACAGGAAACGCGTTGTGCTCTGGCTCCTGGGCTGCCGATGCGCAGGAACGCCTTCTCCATCTACATCCTCAACTTGGCCGCAGCAGACTTCCTCTTCCTCAGCGCGCGCCTTATATATTCCCTGTTAAGCTTCATCAGTATCCCCATACCATCTCTAAAATCCTCTATCCTGTGATGATGTTTTCCTACTTTGCAGGCCTGAGCTCTCTGAGTGCCGTGAGCACCAGCGCTGCCTGTCCGTCTGTGGCCCATCTGGTACCGCTGCCACCGCCCCACACACCTGTGAGCGGTGGTGTGTCTCTGTCTGGCCCTGTCCCTGCTGCGGAGCATCCTGGAGTGGATGTTATGTGGCTTTCTGTTCAGTGGTGCTGATCTGCTTGGTGTCAAACATCAGATTTTCATCACAGTCGCGTGGCTGATTTTATGTGTGGTTCTCTGTGGGTCCA GCCTGGTCTGTGATCAGGATTCTCTGTGGATCCCAGGATACCGCTGACAGGCTGTACGTGACCATCCCGCTCACAGTACTGGTCTTCCTCCTCTGTGGCTGCCCTTTGGCATTTCAGTTTTCCTATTTTATGGATCCACGTGGACAGGGAAGTCTTATTTTGTTCATGTTTCATCTAGTTTCTATTTTCCTGTCCGCTCTAACAGCAGTGCCAACCCATCATTTACTTCTTCGTGGGCTCCTTTAGGCAGCGTCAAAATAGGCAGAACCTGAAGCTGGTTCTCAGAGGGCTCTGCAGGACGCGTCTGAGGTGGATGAAGGTGGAGGCGAGCTTCCTGAGGAAATCCTGGAGCTGTCGGAAGCAGATTGGAGCAGTGAGG

In a search of public sequence databases, the NOV46b nucleic acid sequence, located on chromosome 11, has 494 of 800 bases (61%) identical to a gb:GENBANK-ID:AF295365|acc:AF295365.1 mRNA from *Mus musculus* (G-protein coupled receptor GPR90 mRNA, complete cds) ($E = 1.2e^{-28}$).

The disclosed NOV46b polypeptide (SEQ ID NO:174) encoded by SEQ ID NO:173 has 319 amino acid residues and is presented in Table 46D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV46b has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV46b may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a

certainty of 0.3000. The most likely cleavage site for NOV46b is between positions 42 and 43: TGN-AV.

Table 46D. Encoded NOV46b protein sequence (SEQ ID NO:174).

TIISTLDTELTPTINGTEETLCYKQTLSTVLVTCIVSLVGLTGNVVLWLLGCRMRRNAFSYIILNLAAADFLF
LSGRLIYSLLSFISIPHTISKILYPVMMFSYFAGLSSLSAVSTERCLSVLWPIWYRCHRPHTLSAVVCVLLW
ALSLLRSILEWMLCGFLFSGADSAWCQTSDFITVAWLIFLCVVLCGSSVLVLLIRILCGSRKIPLTRLYVTIP
LTVLVFLLCGLPFGIQFFLFLWIHVDREVLFCVHLVSIPLSALNSSANPIIYFFVGSFRQRQNRQNLKVL
QRALQDASEVDEGGQLPEEILELSGSRLEQ

A search of sequence databases reveals that the NOV46b amino acid sequence has 110 of 275 amino acid residues (40%) identical to, and 174 of 275 amino acid residues (63%) similar to, the 324 amino acid residue ptnr:SWISSPROT-ACC:P12526 protein from *Rattus norvegicus* (Rat) (Mas Proto-Oncogene) ($E = 6.8e^{-45}$).

NOV46c

A disclosed NOV46c nucleic acid of 1030 nucleotides (also referred to as CG56698-01) encoding a Mas Proto-Oncogene-like protein is shown in Table 46E. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 17-19 and ending with a TGA codon at nucleotides 1007-1009. The start and stop codons are shown in bold in Table 46E, and the 5' and 3' untranslated regions, if any, are underlined.

Table 46E. NOV46c nucleotide sequence (SEQ ID NO:175).

ACTAGGGTTTCTGAGCATGAATCCAACCATCCAGCCTTGGATACAGAAATTGCACCAATTAGTGATACAGA
GGAGACCCATCCTCATCGTTGTGGCATGGAGGTCTGGTCCCTCATAGTGCTGATCCTCATCATTGACCTGGT
CGGGCTGGCAGGAAATGCAGTCATGCTCTGGCTCCTGGGCTTCTGCATGCACAGTAACACCTTCTCTCTA
CATCCTCAACCTGGCCAGGGCTGACTTCTCTGCACCTGCTTCCAGATTATAACATTTCATTAATTCTTCAG
TGACTTTGTAGTTCTCTCTCCATCCATTCTCTAGATTGTGACACACGGTGTGTCTCCGCTGTATTAC
AGGCCTGAGCATGCTGAGCACCATCAGCACCAGCACCCTGTCCGTCTGTGGCCCATCTGGTACTGCTG
CCACTGCCCCACACACCTGTGACGGTCTGTGTCTGCTCTGGGCCCTGTCCCTGTGTCAGAGCATCCT
GGAGTGGATGTTCTGTAGCTTCTGTTTAGTGATGTTGACTCTGATAATTGGTGTCAAATATTAGATTTCCT
CACTGCTGTGTGGCTGATTTTTTATCTGTGGTTCTCTGTGGGTTCAACCCTGGTCTGCTGTGTCAGGATCAT
ATGTGGATCCCAGAAGATGCCGCTGACCAGGCTGTATGTGACCATCCTGCTCACAGGGCTGGTCTTCTCTCT
CTGCAGCCTGCCCTCAGCATTAGGGATTCTATTATACTGGATCGAGAAGGATTGGATGACTTACCTTG
TGTTGTTCTGTTTAAATTCATTTTCTGTCTGCTCTTAACAGCAGTGCCAACCCCATCATTTACTTCTTCAT
GGGCTCCTTTAGGCAGCTTCAAAACAGGAAGACCTCAAGCTGGTTCTCCAGAGGGCTCTGCAGGACATGCT
TGAGGTGGATGAAGGTGGAGGGCAGCTTCTGAGGAAACCTGAAGCTGTCCGGAAGCAGATTGGGGCCATG
AGGAAGAGCCTCTGCCCTGTTA

In a search of public sequence databases, the NOV46c nucleic acid sequence, located on chromosome 7, has 381 of 621 bases (61%) identical to a gb:GENBANK-ID:RATMAS|acc:J03823.1 mRNA from *Rattus norvegicus* (Rat mas oncogene, complete cds) ($E = 9.3e^{-22}$).

The disclosed NOV46D polypeptide (SEQ ID NO:176) encoded by SEQ ID NO:175 has 330 amino acid residues and is presented in Table 46F using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV46c has a signal peptide and

is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV46c may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV46c is between positions 65 and 66:
TFS-LY.

Table 46F. Encoded NOV46c protein sequence (SEQ ID NO:176).

MNPTIPALDTEIAPISDTEETHPHRCGMEVLVLIVLILIIDLVGLAGNAVMLWLLGFCMHSNTFSLYILNLA
RADFLCTCFQIITFINFFSDFVSSLSIHFSRFVTTVLFSACITGLSMLSTISTEHLRLSVLWPIWYCCCHCPH
LSAVMCVLLWALSLLQSIIEWMFCSFLFSDVSDNWCQILDFLTAVWLIIFLSVVLGFTLVLLVRIICGSQK
MPLTRLYVTILLTGLVFLFCSLPLSIQGFLLYWIEKDLDDLPCVVRLISIFLSALNSSANPIIYFFMGFSFRQ
LQNRKTLKLVLRALQDMLEVDEGGGQLPEETLKLSGSRLGP

A search of sequence databases reveals that the NOV46c amino acid sequence has 106 of 279 amino acid residues (37%) identical to, and 166 of 279 amino acid residues (59%) similar to, the 324 amino acid residue ptnr:SWISSPROT-ACC:P12526 protein from *Rattus norvegicus* (Rat) (Mas Proto-Oncogene) ($E = 1.6e^{-45}$).

NOV46c is predicted to be expressed in at least teratocarcinoma cell. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in hippocampus and brain because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:RATMAS|acc:J03823.1) a closely related Rat mas oncogene, complete cds homolog.

NOV46d

A disclosed NOV46d nucleic acid of 1005 nucleotides (also referred to as CG56702-01) encoding a Mas Proto-Oncogene-like protein is shown in Table 46G. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 17-19 and ending with a TGA codon at nucleotides 986-988. The start and stop codons are shown in bold in Table 46G, and the 5' and 3' untranslated regions, if any, are underlined.

Table 46G. NOV46d nucleotide sequence (SEQ ID NO:177).

ACTACCGTTTCTGAGCATGGATCCAAGCAACCCAGCCTTGGATGCAGAACTGACACCAATTAACAGAACTGA
GGAGACTCCTTGCTACAAGCAGACCCTGAGCCTCATGGGGCTGACGTGCATCATTTCCCTTGTCACGCTGAC
AGGAAACGCGGTTGTGCTCTGGCTCCTGGGCTTCCGCATGCGCAGGAACGCGTCTCCATCTACATCCTCAA
CCTGGCTGCGGCAGACTTCTCTTCTCTCAGCGGCCACGTTATACATTCCGCCTCACTCCTCATCAATATCTG
TCATCCCATCTCCAAAATCCTCATTCCTGTGATGACCTTCTATACTTTACAGGCCTGAGCTTTCTGAGTGC
CATGAGCACCGAGCGCTGCCTGTGCGTCCCTGTGGCCCATCTGGTACCGCTGCCTCCTCCCCACACACCTGTC
AGCGGTCGTGTGTGCTTGTCTTGGGCCCTGTCCCTACTGCGGAGCATCCTGGAGGGAATGTTCTGTGACTT
CCTGTTTAGTGATGCTGATTCTATTTGGTGTCAACCATCAGATTTTCATCACAGTCGTGTGGCTGATTTTTT
ATGTGTGGTTCTCTGTGGGTCCAGCCTGGTCTGCTGATTAGGATTCTCTGTGGATCCTGGAAGATGCCTCT
GACCGGCTGTACGTGACGATCCTGCTCAGAGTGTAGTCTTCTACTCCGAGCCTGCCCTTCGGCATTCCG
GTGGGCTCTGTCTACTGGGATCCACCTGGATTGGAGTCAATTTCTGTTCATGTCCATCTAGTTTCCATTTT

CCTGTCCCCTCTAAACGGTAGTGCCAAACCCCGTCATTTACTTCTTCGTGGGCTCCTTTAGGCAGCGTCAAAA
TAGGCAGAACCTGAAGCTGGTTCTCCAGAGGGCTCTGCAGGACATGCCTGAGGTGAAGGTGGAAGGTGGAGG
GCGGCTTCCTGAGGGAACCTGGAGCTGTTCGGGAAGCAGATTTCGGGCAGTGAGGAAGAACCTCTGCCCT

In a search of public sequence databases, the NOV46d nucleic acid sequence, located on chromosome 7, has 379 of 632 bases (59%) identical to a gb:GENBANK-ID:RATMAS|acc:J03823.1 mRNA from *Rattus norvegicus* (Rat mas oncogene, complete cds) (E = 4.7e⁻¹⁴).

The disclosed NOV46d polypeptide (SEQ ID NO:178) encoded by SEQ ID NO:177 has 323 amino acid residues and is presented in Table 46H using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV46C has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6000. Alternatively, NOV46d may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV46d is between positions 46 and 47: GNA-VV.

Table 46H. Encoded NOV46d protein sequence (SEQ ID NO:178).

MDPSNPALDAELTPINRTEETPCYKQTLISLMGLTCIISLVLTGNAVVLWLLGFRMRRNAVSIIYILNLAAD
FLFLSGHVIHSASLLINICHPIISKILIPVMTFLYFTGLSFLSAMSTERCLCVLWPIWYRCLLPHTLSAVVCV
LLWALSLLRSILEGMFCDFLFSADSIWCQPSDFITVVWLIPLCVVLCGSSLVLLIRILCGSWKMPITGLYV
TILLTVLVFLLRSLPFGIRWALSTGIHLDLEIVFCHVHLVSIPLSPLNGSANPVIYFFVGSFRQRQNRQNLK
LVLQRALQDMPEVKVEGGGRLPEGTLELSGSRFGQ

A search of sequence databases reveals that the NOV46d amino acid sequence has 107 of 275 amino acid residues (38%) identical to, and 167 of 275 amino acid residues (60%) similar to, the 324 amino acid residue ptnr:SWISSPROT-ACC:P12526 protein from *Rattus norvegicus* (Rat) (Mas Proto-Oncogene) (E = 1.5e⁻⁴⁰).

In addition, NOV46d is predicted to be expressed in hippocampus and cerebral cortex of the brain because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:RATMAS|acc:J03823.1) a closely related Rat mas oncogene, complete cds homolog.

NOV46a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 46I.

Table 46I. BLAST results for NOV46a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17472340 ref XP_061650.1 (XM_061650)	similar to MrgX1 G protein-coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	1589	322/322 (100%)	322/322 (100%)	e-172
gi 15546062 gb AAK91804.1 (AY042213)	MrgX1 G protein-coupled receptor [<i>Homo sapiens</i>]	322	322/322 (100%)	322/322 (100%)	e-158
gi 16876453 ref NP_473372.1 (NM_054031)	G protein-coupled receptor MRGX3 [<i>Homo sapiens</i>]	322	269/322 (83%)	285/322 (87%)	e-131
gi 17461239 ref XP_062249.1 (XM_062249)	similar to MrgX3 G protein-coupled receptor (H. sapiens) [<i>Homo sapiens</i>]	322	268/322 (83%)	285/322 (88%)	e-131
gi 16876455 ref NP_473373.1 (NM_054032)	G protein-coupled receptor MRGX4 [<i>Homo sapiens</i>]	322	255/320 (79%)	275/320 (85%)	e-119

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 46J. In the ClustalW alignment of the NOV46 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 46J. ClustalW Analysis of NOV46

10

Table 46J. ClustalW Analysis of NOV46

15

20

25

30

1) Novel NOV46a (SEQ ID NO:172)

2) Novel NOV46b (SEQ ID NO:174)

3) Novel NOV46c (SEQ ID NO:176)

4) Novel NOV46d (SEQ ID NO:178)

5) gi|17472340|ref|XP_061650.1| (XM_061650) similar to MrgX1 G protein-coupled receptor (H. sapiens) [Homo sapiens] (SEQ ID NO:532)

6) gi|15546062|gb|AAK91804.1| (AY042213) MrgX1 G protein-coupled receptor [Homo sapiens] (SEQ ID NO:533)

7) gi|16876453|ref|NP_473372.1| (NM_054031) G protein-coupled receptor MRGX3 [Homo sapiens] (SEQ ID NO:534)

8) gi|17461239|ref|XP_062249.1| (XM_062249) similar to MrgX3 G protein-coupled receptor (H. sapiens) [Homo sapiens] (SEQ ID NO:535)

9) gi|16876455|ref|NP_473373.1| (NM_054032) G protein-coupled receptor MRGX4 [Homo sapiens] (SEQ ID NO:536)

10

20

30

40

50

60

NOV46a 1|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

5	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
10	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
15	gi 17472340	61	PINRTEETPCYKQTLNLMGLTCIISLVTLTGNAVVLWLLGFRMRNNAVSIYILNLAAADF	120
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
20	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
25	gi 17472340	121	LFLSGHVIHSASLLINICHPIISKILIPVMTFLYFTGLSFLSAMSTERCLCVLPICLVLL	180
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
30	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
35	gi 17472340	181	IRILCGSWKMPLTGLYVTILLTVLVFLRLSLPFGIRWALSTGIHLDLEVI FCHVHLVSIF	240
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
40	gi 16876455	1	-----	1
45	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
50	gi 17472340	241	LSPLNGSANPVIYFFVGSRQRQNRQNLKLVLRALQDMPEVKVEGGFLREPWSCREADS	300
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
55	gi 16876455	1	-----	1
	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
60	NOV46d	1	-----	1
	gi 17472340	301	GTVTAEYRRRNLAHAHHIGPNFSETLQDENTRKESRDRADLIPECASPELSNTEESSEV	360
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
65	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
70	NOV46a	1	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1

5	NOV46d	1	-----	1
	gi 17472340	361	TESQKGGPPPLAQGKSSPLDKREKDKGKYQELSSSIVSYGFCYKIVSVFPPrDTRLGFLSV	420
	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
10	NOV46a	1	430 440 450 460 470 480	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	421	DPTIPALGTELTPIINGREETPCYKQTLSTLTGLTCIVSLVGMTGNAVVLWLLGFRMRNAF	480
15	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
20	NOV46a	1	490 500 510 520 530 540	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	481	SIYIFNLSMADFLFLRSHIIRFPLSLINILHPIFKILSPVMMFSYLASLSFLSAMSTERC	540
25	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
30	NOV46a	1	550 560 570 580 590 600	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	541	LLVLLVRILCGSRKMPPLRLQCNQNLKLVLRALQDTTEVNEGGRWLPEETLELSGSR	600
35	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
40	NOV46a	1	610 620 630 640 650 660	1
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	601	FGQSVHCLWEEEWRCGSLMSCLLSKVNDITSSWFHIKEEHAGHPGVFSRKVTRLGFLS	660
45	gi 15546062	1	-----	1
	gi 16876453	1	-----	1
	gi 17461239	1	-----	1
	gi 16876455	1	-----	1
50	NOV46a	1	670 680 690 700 710 720	1
	NOV46b	1	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	NOV46c	1	---TISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	54
	NOV46d	1	MNPTIPALDTEIAPISDTEETHPHRCGMEVLVILVILILIDLVGLAGNAVVLWLLGFCMH	60
	gi 17472340	661	MDPSNPALDAELTPINRTEETP---CYKQTLSTLMGLTCTISLVLTGNNAVVLWLLGFRMR	57
55	gi 15546062	1	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	717
	gi 16876453	1	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 17461239	1	MDSTIPVLGTELTPIINGREETP---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 16876455	1	MDSTIPVLGTELTPIINGREETP---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 16876455	1	MDPTIPVFGTILTPINGREETP---CYNQTLSTLTVCITISLVGLTGNNAVVLWLLGFRMR	57
60	NOV46a	1	730 740 750 760 770 780	1
	NOV46b	1	RNAFSIYILNLAAADFLFLSGLIYSL---LSFTIS-TPHTISKILYVPMMFSSYFAGLSFL	113
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	661	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
65	gi 15546062	1	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 16876453	1	MDSTIPVLGTELTPIINGREETP---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 17461239	1	MDSTIPVLGTELTPIINGREETP---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 16876455	1	MDSTIPVLGTELTPIINGREETP---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57
	gi 16876455	1	MDPTIPVFGTILTPINGREETP---CYNQTLSTLTVCITISLVGLTGNNAVVLWLLGFRMR	57
70	NOV46a	58	RNAFSIYILNLAAADFLFLSGLIYSL---LSFTIS-TPHTISKILYVPMMFSSYFAGLSFL	113
	NOV46b	1	-----	1
	NOV46c	1	-----	1
	NOV46d	1	-----	1
	gi 17472340	661	MDPTISTLDTELTPIINGTEETL---CYKQTLSTLTVCIVSLVGLTGNNAVVLWLLGCRMR	57

5	NOV46b	55	RNAFSIYILNLAADFLFLSGRLIYSL---LSFIS-IPHTISKILYPVMMFSYFAGLSL	110
	NOV46c	61	SNIFSLYILNLAADFLCTCFQIITFINFFSDFVSSSIHFSRFVTIVIFSACITGLSML	120
	NOV46d	58	RNAVSIYILNLAADFLFLSGHVTHSA---SLLIN-ICHPISKILYPVMTFFLYETGLSFL	113
	gi 17472340	718	RNAFSIYILNLAADFLFLSGRLIYSL---LSFIS-IPHTISKILYPVMMFSYFAGLSFL	773
	gi 15546062	58	RNAFSIYILNLAADFLFLSGRLIYSL---LSFIS-IPHTISKILYPVMMFSYFAGLSFL	113
10	gi 16876453	58	RNAVSIYILNLAADFLFLSGHITCSP---LRLIN-IRHPISKILSPVMTFFPYETGLSML	113
	gi 17461239	58	RNAVSIYILNLAADFLFLSGHITCSP---LRLIN-IRHPISKILSPVMTFFPYETGLSML	113
	gi 16876455	58	RNAVSIYILNLAADFLFLSQITRSP---LRLIN-ISELIRKILVSVMTFFPYETGLSML	113
15	NOV46a	114	SAVSTERCLSVLWPIWYRCHRPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSGADS-AW	172
	NOV46b	111	SAVSTERCLSVLWPIWYRCHRPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSGADS-AW	169
	NOV46c	121	STLSTEHRLSVLWPIWYCHCPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSDVDSNW	180
	NOV46d	114	SAMSTERCLSVLWPIWYRCLLPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSDADS-IW	172
	gi 17472340	774	SAVSTERCLSVLWPIWYRCHRPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSGADS-AW	832
20	gi 15546062	114	SAVSTERCLSVLWPIWYRCHRPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSGADS-AW	172
	gi 16876453	114	SAMSTERCLSVLWPIWYHCRRPRLSSVMCVLLWALSLLRSILEWMLCGFLFSGADS-VW	172
	gi 17461239	114	SAMSTERCLSVLWPIWYHCRRPRLSSVMCVLLWALSLLRSILEWMLCGFLFSGADS-VW	172
	gi 16876455	114	SAMSTERCLSVLWPIWYRCHRPHTLSAVVCVLLWALSLLRSILEWMLCGFLFSGADS-SW	172
25	NOV46a	173	CQTSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
	NOV46b	170	CQTSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	229
	NOV46c	181	CQILDFTVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	240
	NOV46d	173	CQPSDFITVWVLIFFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
	gi 17472340	833	CQTSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	892
30	gi 15546062	173	CQTSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
	gi 16876453	173	CETSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
	gi 17461239	173	CETSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
	gi 16876455	173	CETSDFITVAWLIFLCVVLGSSLVLLIRILCGSRKIPLTRLYVTILLTVLVFLLCGLPF	232
35	NOV46a	233	GIOFFLELWIHVDREVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
	NOV46b	230	GIOFFLELWIHVDREVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	289
	NOV46c	241	SIQGFLLYWIEKDDLLPCVRLIISIFLSALNSSANPIIYFFVGSFRQLQNRKTLKVLQ	300
	NOV46d	233	GIRWALSTGIHLDLEVLFCVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
	gi 17472340	893	GIOFFLELWIHVDREVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	952
40	gi 15546062	233	GIOFFLELWIHVDREVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
	gi 16876453	233	GIOWALSRIHLDWKVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
	gi 17461239	233	GIOWALSRIHLDWKVLFCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
	gi 16876455	233	GILGALTIYRMHLENLEVLVCHVHLVSIIFLSALNSSANPIIYFFVGSFRQRQRNQLKVLQ	292
45	NOV46a	293	RALQDASEVD-EGGGQLPEETLELSGSRLEQ-----	322
	NOV46b	290	RALQDASEVD-EGGGQLPEETLELSGSRLEQ-----	319
	NOV46c	301	RALQDMLEVD-EGGGQLPEETLELSGSRLEQ-----	330
	NOV46d	293	RALQDMPEVKVEGGGRLPEETLELSGSRLEQ-----	323
	gi 17472340	953	RALQDASEVD-EGGGQLPEETLELSGSRLEQGTRLGFLSMDPTIPVLGTETLPINGTEET	1011
50	gi 15546062	293	RALQDASEVD-EGGGQLPEETLELSGSRLEQ-----	322
	gi 16876453	293	RALQDTPEVD-EGGGWLPQETLELSGSRLEQ-----	322
	gi 17461239	293	RALQDTPEVD-EGGGWLPQETLELSGSRLEQ-----	322
	gi 16876455	293	RALQDKPEVD-KCGGQLPEETLELSGSRLEQ-----	322
55	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
	NOV46d	323	-----	323
	gi 17472340	1012	PCYNQTLSTFTVLTICIVSLVALTGNAVVLLGFRMCRNAVSIYILNLAANFLLSSHII	1071
60	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
	gi 16876455	322	-----	322
65	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
	NOV46d	323	-----	323
	gi 17472340	1012	PCYNQTLSTFTVLTICIVSLVALTGNAVVLLGFRMCRNAVSIYILNLAANFLLSSHII	1071
70	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
	gi 16876455	322	-----	322

			
	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
5	NOV46d	323	-----	323
	gi 17472340	1072	HPCYTSSITPEYSECHEHQALPVNPVAHLPGPSGQDPLWIPEDAADQAVHDHLLTVLVFL	1131
	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
10	gi 16876455	322	-----	322
			1150 1160 1170 1180 1190 1200	
			
	NOV46a	322	-----	322
15	NOV46b	319	-----	319
	NOV46c	330	-----	330
	NOV46d	323	-----	323
	gi 17472340	1132	LCGLPIGIQWALFSRIHMDWEVLYSHVHLPSIFLSSLNSSANPIIYFPMGFVRQHQNWN	1191
	gi 15546062	322	-----	322
20	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
	gi 16876455	322	-----	322
			1210 1220 1230 1240 1250 1260	
			
	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
	NOV46d	323	-----	323
30	gi 17472340	1192	LKLVLRDLQDTPEVDEEQNNVLRYYIILYIMYTLDTMSLHKNSLGNWFGKSSAVICLQN	1251
	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
	gi 16876455	322	-----	322
35			1270 1280 1290 1300 1310 1320	
			
	NOV46a	322	-----	322
	NOV46b	319	-----	319
40	NOV46c	330	-----	330
	NOV46d	323	-----	323
	gi 17472340	1252	GDMALVSSKDNHGPFLRTDIGGTGLLFLAVTAEYRRRNLNAAHGHGPIVLKPCRMT	1311
	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
45	gi 17461239	322	-----	322
	gi 16876455	322	-----	322
			1330 1340 1350 1360 1370 1380	
			
	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
	NOV46d	323	-----	323
50	gi 17472340	1312	QGRREKTQWISFLSVYVEISPKLHNTKGSSEVTESQKGSTSLARGSTSSTLDRRRRKDAQ	1371
	gi 15546062	322	-----	322
55	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
	gi 16876455	322	-----	322
			1390 1400 1410 1420 1430 1440	
			
	NOV46a	322	-----	322
	NOV46b	319	-----	319
	NOV46c	330	-----	330
60	NOV46d	323	-----	323
	gi 17472340	1372	QSHIEPHFKGTLVVNLRGTRLGFLSMNPTIPALDTEIAPISDTEETHPHRCGMEVLVLIV	1431
	gi 15546062	322	-----	322
	gi 16876453	322	-----	322
	gi 17461239	322	-----	322
70	gi 16876455	322	-----	322

			1450	1460	1470	1480	1490	1500	
5	NOV46a	322	322
	NOV46b	319	-----	-----	-----	-----	-----	-----	319
	NOV46c	330	-----	-----	-----	-----	-----	-----	330
	NOV46d	323	-----	-----	-----	-----	-----	-----	323
	gi 17472340	1432	LILIIDLVLGLAGNAVMLWLLGFCMHSNTFSLYILNLARADFLCTCFQIITFINFFSDFVS						1491
10	gi 15546062	322	-----	-----	-----	-----	-----	-----	322
	gi 16876453	322	-----	-----	-----	-----	-----	-----	322
	gi 17461239	322	-----	-----	-----	-----	-----	-----	322
	gi 16876455	322	-----	-----	-----	-----	-----	-----	322
15	NOV46a	322	322
	NOV46b	319	-----	-----	-----	-----	-----	-----	319
	NOV46c	330	-----	-----	-----	-----	-----	-----	330
	NOV46d	323	-----	-----	-----	-----	-----	-----	323
20	gi 17472340	1492	SLSIHFSRFVTTVLFSACITGLSMLSTISTEHRSLVLPICSANPIIYFFMGFSRQLQNR						1551
	gi 15546062	322	-----	-----	-----	-----	-----	-----	322
	gi 16876453	322	-----	-----	-----	-----	-----	-----	322
	gi 17461239	322	-----	-----	-----	-----	-----	-----	322
	gi 16876455	322	-----	-----	-----	-----	-----	-----	322
25	NOV46a	322	322
	NOV46b	319	-----	-----	-----	-----	-----	-----	319
	NOV46c	330	-----	-----	-----	-----	-----	-----	330
	NOV46d	323	-----	-----	-----	-----	-----	-----	323
30	gi 17472340	1552	KTLKLVLRALQDMLEVDEGGQLPEETLKLSGSRLGP						1589
	gi 15546062	322	-----	-----	-----	-----	-----	-----	322
	gi 16876453	322	-----	-----	-----	-----	-----	-----	322
35	gi 17461239	322	-----	-----	-----	-----	-----	-----	322
	gi 16876455	322	-----	-----	-----	-----	-----	-----	322

Table 46K lists the domain description from DOMAIN analysis results against NOV46. This indicates that the NOV46 sequence has properties similar to those of other proteins known to contain this domain.

Table 46K Domain Analysis of NOV46	
gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)	
CD-Length = 254 residues, 37.8% aligned	
Score = 51.2 bits (121), Expect = 9e-08	

45	NOV53: 44	GNAVVLWLLGCRMR-RNAFSIYILNLAAADFLFLSGRLIYSLLSFIS----	IPHTISKIL 98
	Sbjct: 1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
	NOV53: 99	YPVMMFSYFAGLSFSLSAVSTERCLSVLWPIWYRCHR	134
50	Sbjct: 61	GALFVVNGYASILLTIAISIDRYLAIVHPLRYRRIR	96

The Mas Proto Oncogene belongs to the family of G-Protein Coupled Receptors. G-protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide

range of functions (including various autocrine, paracrine and endocrine processes). They show considerable diversity at the sequence level, on the basis of which they can be separated into distinct groups. The currently known clan members include the rhodopsin-like GPCRs, the secretin-like GPCRs, the cAMP receptors, the fungal mating pheromone receptors, and the metabotropic glutamate receptor family.

The rhodopsin-like GPCRs themselves represent a widespread protein family that includes hormone, neurotransmitter and light receptors, all of which transduce extracellular signals through interaction with guanine nucleotide-binding (G) proteins. Although their activating ligands vary widely in structure and character, the amino acid sequences of the receptors are very similar and are believed to adopt a common structural framework comprising 7 transmembrane (TM) helices. The human mas oncogene was originally detected by its ability to transform NIH 3T3 cells. We previously showed that the protein encoded by this gene is unique among cellular oncogene products in that it has seven hydrophobic potential transmembrane domains and shares strong sequence similarity with a family of hormone-receptor proteins. We have now cloned the rat homolog of the mas oncogene, determined its DNA sequence, and examined its expression in various rat tissues. A comparison of the predicted sequences of the rat and human mas proteins shows that they are highly conserved, except in their hydrophilic amino-terminal domains. Our examination of the expression of mas, determined by RNA-protection studies, indicates that high levels of mas RNA transcripts are present in the hippocampus and cerebral cortex of the brain, but not in other neural regions or in other tissues. This pattern of expression and the similarity of mas protein to known receptor proteins suggest that mas encodes a receptor that is involved in the normal neurophysiology and/or development of specific neural tissues.

The disclosed NOV46 nucleic acid of the invention encoding a Mas Proto-Oncogene-like protein includes the nucleic acid whose sequence is provided in Table 46A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 46A while still encoding a protein that maintains its Mas Proto-Oncogene-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or

derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 40 percent of the bases may be so changed.

5 The disclosed NOV46 protein of the invention includes the Mas Proto-Oncogene-like protein whose sequence is provided in Table 46B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 46B while still encoding a protein that maintains its Mas Proto-Oncogene-like activities and physiological functions, or a functional fragment thereof. In the mutant or
10 variant protein, up to about 21 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above disclosed information suggests that this Mas Proto-Oncogene -like protein (NOV46) is a member of a "Mas Proto-Oncogene family". Therefore, the NOV46 nucleic
15 acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug
20 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV46 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease,
25 Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, and/or other diseases and pathologies.

 NOV46 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV46 substances for use in
30 therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV46 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in

understanding of pathology of the disease and development of new drug targets for various disorders.

NOV47

A disclosed NOV47 nucleic acid of 523 nucleotides (also referred to as AF152363) encoding a Peptidyl-Prolyl Cis-Trans Isomerase -like protein is shown in Table 47A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 17-19 and ending with a TAA codon at nucleotides 509-511. The start and stop codons are shown in bold in Table 47A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 47A. NOV47 nucleotide sequence (SEQ ID NO:179).

```
CCCCGTATTACCAGCTATGGTCAACCCCACTGTTTCTTCGACATTGCTGTCAATAGCGAGCCCTTGGGCTG
CGTCTCCTTCGAGCTGTTTGCAGACAAGCTTCCAAAGACAGCAGAAAATTTTCATGCTCTGAGCACTGGAGA
AAAAGGATTTGATTATGAGGGTTACTGCTTTACAGAATTATTCAGGGTTTGTATGTCAGGGTGGTGACTT
CACATGCCATAATGGCACTGGTAGCAAGTCCATCTACAGGGAGAAATTTGATGACGAGAACTTCATCCTGAA
GCATACAGGTCCTGGCATCCTGTCCATGGCAAAATGCTGGACCCAACGCAAATGGTTCCAGTTTTCATGTG
CCCTGCCAAGACCAAGTGGTTGGATGGCAAGCAAGTGGTCTTTGGCAGGGTGAAAGAAGGCATGGATATGT
GGAGGCCATGGAGCGCTTTGTGTTTCAGGAATGGCAAGACTAGCAAGAAGGTCACTATTGCTGACTGTGGACA
GCTCTAATAAGTTTGACTT
```

In a search of public sequence databases, the NOV47 nucleic acid sequence, located on chromosome 3, has 523 of 523 bases (100%) identical to a gb:GENBANK-ID:AF152363|acc:AF152363.1 mRNA from *Homo sapiens* (constitutive fragile region FRA3B sequence) ($E = 5.7e^{-111}$).

The disclosed NOV47 polypeptide (SEQ ID NO:180) encoded by SEQ ID NO:179 has 164 amino acid residues and is presented in Table 47B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV47 has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.6400. Alternatively, NOV47 may also localize to the plasma membrane with a certainty of 0.6000, to the cytoplasm with a certainty of 0.4500, or to the mitochondrial matrix space with a certainty of 0.1000.

Table 47B. Encoded NOV47 protein sequence (SEQ ID NO:180).

```
MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEYGYCFHRIIPGFVCQGGDFTCHNG
TGSKSIYREKFDENFILKHTGPGILSMANAGPNANGSQFFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMER
FVFRNGKTSKKVTIADCGQL
```

A search of sequence databases reveals that the NOV47 amino acid sequence has 141 of 164 amino acid residues (85%) identical to, and 151 of 164 amino acid residues (92%) similar to, the 165 amino acid residue ptrn:pir-id:CSHUA protein from human (peptidylprolyl isomerase (EC 5.2.1.8) A) ($E = 5.6e^{-75}$).

NOV47 is predicted to be expressed in small intestine because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:E02765|acc: E02765.1) a closely related *Sus scrofa* peptidyl-prolyl cis-trans isomerase A sequence homolog.

NOV47 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 47C.

Table 47C. BLAST results for NOV47					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17440554 ref XP_067503.1 (XM_067503)	similar to PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A(PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A- BINDING PROTEIN) (H. sapiens) [Homo sapiens]	164	164/164 (100%)	164/164 (100%)	3e-91
gi 12804335 gb AAH03026.1 AAH03026 (BC003026)	Unknown (protein for IMAGE:2823490) [Homo sapiens]	174	141/164 (85%)	151/164 (91%)	3e-76
gi 4033689 sp P04374 CYPH_BOVIN	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOP HILIN A) (CYCLOSPORIN A- BINDING PROTEIN)	164	141/164 (85%)	151/164 (91%)	1e-75
gi 10863927 ref NP_066953.1 (NM_021130)	peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens]	165	141/164 (85%)	151/164 (91%)	2e-75
gi 68401 pir CSBOA B	peptidylprolyl isomerase (EC 5.2.1.8) A - bovine	163	140/163 (85%)	150/163 (91%)	5e-75

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 47D. In the ClustalW alignment of the NOV47 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 47D. ClustalW Analysis of NOV47

1) Novel NOV47 (SEQ ID NO:180)
 2) gi|17440554|ref|XP_067503.1| (XM_067503) similar to PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A-BINDING PROTEIN) (H. sapiens) [Homo sapiens] (SEQ ID NO:537)
 3) gi|12804335|gb|AAH03026.1|AAH03026 (BC003026) Unknown (protein for IMAGE:2823490) [Homo sapiens] (SEQ ID NO:538)
 4) gi|4033689|sp|P04374|CYPH_BOVIN PEPTIDYL-PROLYL CIS-TRANS ISOMERASE A (PPIASE) (ROTAMASE) (CYCLOPHILIN A) (CYCLOSPORIN A-BINDING PROTEIN) (SEQ ID NO:539)
 5) gi|10863927|ref|NP_066953.1| (NM_021130) peptidylprolyl isomerase A (cyclophilin A) [Homo sapiens] (SEQ ID NO:540)
 6) gi|68401|pir|CSBOAB peptidylprolyl isomerase (EC 5.2.1.8) A - bovine (SEQ ID NO:541)

			10	20	30	40	50	60	
NOV47	1							
gi 17440554	1	-----MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGY	51						
gi 12804335	1	-----MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGY	51						
gi 4033689	1	-----MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGY	51						
gi 10863927	1	-----MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGY	51						
gi 68401	1	-----MVNPTVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGY	51						
			70	80	90	100	110	120	
NOV47	52							
gi 17440554	52	CFHRIIPGFVCOGGDFTCHNGTGSKSIYREKFDDENFILKHTGPGILSMANAGPNANGSQ	111						
gi 12804335	61	CFHRIIPGFVCOGGDFTCHNGTGSKSIYREKFDDENFILKHTGPGILSMANAGPNANGSQ	111						
gi 4033689	52	CFHRIIPGFVCOGGDFTCHNGTGSKSIYREKFDDENFILKHTGPGILSMANAGPNANGSQ	111						
gi 10863927	52	CFHRIIPGFVCOGGDFTCHNGTGSKSIYREKFDDENFILKHTGPGILSMANAGPNANGSQ	111						
gi 68401	51	CFHRIIPGFVCOGGDFTCHNGTGSKSIYREKFDDENFILKHTGPGILSMANAGPNANGSQ	110						
			130	140	150	160	170		
NOV47	112							
gi 17440554	112	FFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMERFVERNGKTSKKVTIADCGQL	164						
gi 12804335	121	FFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMERFVERNGKTSKKVTIADCGQL	164						
gi 4033689	112	FFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMERFVERNGKTSKKVTIADCGQL	164						
gi 10863927	112	FFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMERFVERNGKTSKKVTIADCGQL	164						
gi 68401	111	FFMCPAKTKWLDGKQVVFGRVKEGMDIVEAMERFVERNGKTSKKVTIADCGQL	163						

Table 47E lists the domain descriptions from DOMAIN analysis results against NOV47. This indicates that the NOV47 sequence has properties similar to those of other proteins known to contain this domain.

Table 47E Domain Analysis of NOV47

gnl|Pfam|pfam00160, pro_isomerase, Cyclophilin type peptidyl-prolyl cis-trans isomerase (SEQ ID NO:839)
 CD-Length = 162 residues, 100.0% aligned
 Score = 219 bits (558), Expect = 1e-58

NOV47: 5 TVFFDIAVNSEPLGCVSFELFADKLPKTAENFHALSTGEKGFDEGYCFHRIIPGFVCO 63
 Sbjct: 1 KVFYFDITIGGKPLGRIVFELFGDVVPKTVENFRALCTGEKGFDEGYCFHRIIPGFVCO 60

5 NOV47: 64 GGDF TCHNGT GSKSIYREKFD DENFILKHTGPGILSMANAGPNANGSQFFMCPAKTKWLD 123
 Sb jct: 61 GGDFTRGNGTGGKSIYGEKFDENFNLKHDRPGTLSMANAGPNTNGSQFFITTVATPWLD 120

11 NOV47: 124 GKQVVFGRVKEGMDIVEAMERFVFRNG-KTSKKVTIADCGQL 164
 Sb jct: 121 GKHVVFGRVKEGMDVVDKIENVGTDSGDVPSKDVKIADCGQL 162

Cyclophilin A (peptidyl-prolyl cis-trans isomerase A) is the major high-affinity
 10 binding protein in vertebrates for the immunosuppressive drug cyclosporin A (CSA). Because
 of its dramatic effects on decreasing morbidity and increasing survival rates in human
 transplants, the molecular mechanism of immunosuppression by cyclosporin A has been a
 matter of much interest. Cyclophilin A is a member of the immunophilin class of proteins that
 all possess peptidyl-prolyl cis-trans isomerase (PPIase) activity and, therefore, are believed to
 15 be involved in protein folding and/or intracellular protein transport. PPIase accelerates protein
 folding by catalyzing the cis-trans isomerization of proline imidic peptide bonds in
 oligopeptides. It is probable that CSA mediates some of its effects via an inhibitory action on
 PPIase. Cyclophilin is a cytosolic protein that belongs to a family of isozymes, including
 cyclophilins B and C, PPIase, and natural killer cell cyclophilin-related protein. The sequences
 20 of the different forms of cyclophilin-type PPIases are well conserved. Additional interest in
 cyclophilin A stems from studies performed by Luban et al. (1993), who showed that
 cyclophilin A binds to the gag protein of human immunodeficiency virus type 1 (HIV-1). This
 interaction can be inhibited by the immunosuppressant cyclosporin A and also by
 nonimmunosuppressive, cyclophilin A-binding cyclosporin A derivatives, which were also
 25 shown to exhibit potent anti-HIV-1 activity. Thus, cyclophilin A may have an essential
 function in HIV-1 replication.

The disclosed NOV47 nucleic acid of the invention encoding a Peptidyl-Prolyl Cis-
 Trans Isomerase-like protein includes the nucleic acid whose sequence is provided in Table
 47A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of
 30 whose bases may be changed from the corresponding base shown in Table 47A while still
 encoding a protein that maintains its Peptidyl-Prolyl Cis-Trans Isomerase -like activities and
 physiological functions, or a fragment of such a nucleic acid. The invention further includes
 nucleic acids whose sequences are complementary to those just described, including nucleic
 acid fragments that are complementary to any of the nucleic acids just described. The
 35 invention additionally includes nucleic acids or nucleic acid fragments, or complements
 thereto, whose structures include chemical modifications. Such modifications include, by way
 of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones
 are modified or derivatized. These modifications are carried out at least in part to enhance the

chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

5 The disclosed NOV47 protein of the invention includes the Peptidyl-Prolyl Cis-Trans Isomerase-like protein whose sequence is provided in Table 47B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 47B while still encoding a protein that maintains its Peptidyl-Prolyl Cis-Trans Isomerase-like activities and physiological functions, or a functional fragment
10 thereof. In the mutant or variant protein, up to about 15 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

 The above disclosed information suggests that this Peptidyl-Prolyl Cis-Trans
15 Isomerase -like protein (NOV47) is a member of a "Peptidyl-Prolyl Cis-Trans Isomerase family". Therefore, the NOV47 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target
20 (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV47 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in inflammatory bowel disease, diverticular disease, and/or
25 other diseases and pathologies.

 NOV47 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV47 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-
30 NOVX Antibodies" section below. The disclosed NOV47 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV48

NOV48 includes three novel Phospholipase C Delta-4-like proteins disclosed below. The disclosed sequences have been named NOV48a and NOV48b.

NOV48a

- 5 A disclosed NOV48a nucleic acid of 3238 nucleotides (also referred to as CG56743-01) encoding a Phospholipase C Delta-4-like protein is shown in Table 48A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 370-372 and ending with a TGA codon at nucleotides 2626-2628. The start and stop codons are shown in bold in Table 48A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 48A. NOV48a nucleotide sequence (SEQ ID NO:181).

TTCTGATCTTTATGTGATTTGAGGCAGGTTTTCTAAACCTATCTAAAGTGTGAGAGTCACTAAACTCAAATT
AGAAGCAAAATCAGCTACAGACTATCTTCAAGATTACCCAGAGCCCTTTGCTCTTCCTTGCTCCTTTAGG
TGATCTGGTGCCAGCTGGTGGAAACAGTGGGTGATGGCGTCCCTGCTGCAAGACCGTGAGTGCCGGGGCCCT
GCAGGGGAAGAGGCCTTAGTGTACAGCTCAGGGAAGGAGGTTGGACCCCTGTTCCAGAGCTCTCCCT
GGGCTGCTACCCCTCTCTGCTGGCTACCTAACCCTGCTTTTCTGACCTAGAGCTGACCACTGATCAGGAC
TTGCTGCTGATGCAGGAAGGCATGCCGATGCGCAAGGTGAGGTCCAAAGCTGGAAGAAGCTAAGATACTTC
AGACTTCAGAATGACGGCATGACAGTCTGGCATGCACGGCAGGCCAGGGGCAGTGCCAAGCCAGCGTCTCA
ATCTCTGATGTGGAGACAATACGTAATGGCCATGATTCCGAGTTGCTGCGTAGCCTGCCAGAGGAGCTCCCC
CTGGAGCAGGGCTTCACCATTGTCTTCCATGGCCGCCGCTCCAACCTGGACCTGATGGCCAACAGTGTGAG
GAGGCCCAGATATGGATGCGAGGGCTCCAGCTGTGGTGGATCTTGTACCAGCATGGACCATCAGGAGCGC
CTGGACCAGTATCTGAGCGATTGGTTTCAACGTGCAGACAAAAATCAGGATGGTAAGATGAGTTTCCAAGAA
GTCAGCGGTTATTGCACCTAATGAATGTGGAAATGGACCAAGAATATGCCTTCAGTCTCTTGACGGCAGCA
GACACGTCCCAGTCTGGAACCCCTGGAAGGAGAAGAATTCTGACAGTTCTATAAGGCATTGACTAAACGTGCT
GAGGTGCAGGAAGTGTGAAAGTTTTCAGCTGATGGGCAGAGCTGACTCTGCTGGAATTTTGGATTTC
CTCCAAGAGGAGCAGAAGGAGAGAGACTGCACCTTGAGCTTGCTCTGGAACCTATTGACCGCTATGAACCT
TCAGACAGTGGTAAGCTGCGGCATGTGCTGAGTATGGATGGCTTCTCAGCTACCTCTGCTCTAAGGATGGA
GACATCTTCAACCCAGCTGCCTCCCATCTATCAGGATATGACTCAACCCCTGAACCACTACTTCATCTGC
TCTTCTCATAACACCTACCTAGTGGGGGACAGCTTTGTGGCCAGAGCAGCGTCGAGGGATATATACGGGCC
CTGAAGCGGGGTGCCGCTGCGTGGAGGTGGATGTATGGGATGGACCTAGCGGGGAACCTGTCTGTTTACCAC
GGACACACCCCTGACCTCCCGCATCTGTTCAAAGATGTCTGGCCACAGTAGCACAGTATGCCTTCCAGACA
TCAGACTACCCAGTCATCTGTCCCTGGAGACCACTGCAGCTGGGAGCAGCAGCAGACCATGGCCCGTCAT
CTGACTGAGATCCTGGGGGAGCAGCTGCTGAGCACCACCTGGATGGGGTCTGCCCACTCAGCTGCCCTCG
CCTGAGGAGCTTCGGAGGAAGATCTCGTGAAGGGGAAGAAGTTAACACTTGAGGAAGACCTGGAATATGAG
GAAGAGGAAGCAGAACCTGAGTTGGAAGAGTCAGAATTGGCGCTGGAGTCCCAGTTTGAGACTGAGCCTGAG
CCCCAGGAGCAGAACCTTCAGAATAAGGACAAAAGAAGTAAGCCAGCTTCTCCAGAAATCCAAGCCCATC
TTGTGTCCAGCCCTCTCTCCCTGGTTATCTACTTGAAGTCTGTCTCATTCGCGAGCTTCACACATTCAAAG
GAGCACTACCACTTCTACGAGATATCATCTTTCTGAAACCAAGGCCAAGCGCCTCATCAAGGAGGCTGGC
AATGAGTTTGTGCAGCACAATACCTGGCAGTTAAGCCGTGTGTATCCAGCGGCCCTGAGGACAGACTCTTCC
AACTACTACAACCCCAAGAACTCTGGAATGCAGGCTGCCAGATGGTGGCCATGAATATGCAGACTGCAGGG
CTTGAATATGGACATCTGTGATGGGCATTTCCGCCAGAATGGCGGCTGTGGCTATGTGCTGAAGCCAGACTTC
CTGCGTGATATCCAGAGTTCTTTCCACCCTGAGAAGCCCATCAGCCCTTTCAAAGCCAGACTCTCTTAAAC
CAGGTGATCAGCGGTCAGCAACTCCCAAAGTGACAAGACCAAAGAGGGGTCCATGTGGATCCACTGGTG
AAAGTGCAGATCTTTGGCGTTCGTCTAGACACAGCAGGCAGGAGACCAACTATGTGGAGAACAATGGTTTT
AATCCATACTGGGGGCAGACACTATGTTTCCGGGTGCTGGTGCCTGAACTTGCCATGCTGCGTTTGTGGTA
ATGGATTATGACTGGAATCCCGAAATGACTTTATTTGGTCACTACACCTGCCTTGGACCTGCATGCAACAA
GGTTACCGCCACATTACCTGCTGTCAAAGATGGCATCAGCCTCCGCCAGCTTCCATCTTTGTGTATATC
TGCATCCAGGAAGGCCTGGAGGGGGATGAGTCTGAGGTGGGCATTTACGGGAAGGGTTGGTGTGCTGGCT
TTAGAGGTGGAAGCAACATCTGGAAGGATGCTCGAGAGCAACAAATGGAGGTGGTGAAGATTTGGATT
GTGCATTCTTAGGCACAAAATTACCTCATTCTTCTAACAAGCAATCTGGGACCTGATTTTCCACCTTTTTT
CTCTTTCTCTCCTTCTTGTGTTTCTAAGCCTTTGGTATCTTCTCCTGCCCTTTCTCTTGTGTACTCTAT
ACTGGAGTTCCCTTCTTCTCTTGTGCTGATGGCTCAATCCCATACCGACATCTACAATAATCTTCCCATCA
ACTCTGTGTGAAGGAGGTTGCAACTAGAAATTCAGAGGGGCTTGAATAGAGAACCTAAAGAAGCATCAT
CCCCTCCATCCCAACTTCTCAAAGCCCAAGCCAAGGGAAGGATAAATCAAGGCTCAAGGCTTCCCAGC
AAAGATTAGGGAAGAGACTTGACCCAGGACTGTACTACGACTCTTAAGAGAACACTGCACAGCACTCAAA
GTCCCCCACTGGACTGCTTCTCTTAGCCCCACTGGTATAAATACATCTCTCTCAATTTGGCTTCAA

In a search of public sequence databases, the NOV48a nucleic acid sequence, located on chromosome 3, has 1279 of 1285 bases (99%) identical to a gb:GENBANK-ID:AK023083|acc:AK023083.1 mRNA from *Homo sapiens* (cDNA FLJ13021 fis, clone NT2RP3000742, weakly similar to 1-Phosphatidylinositol-4,5-Bisphosphate Phosphodiesterase Delta 1 (EC 3.1.4.11)) (E = 0.0).

The disclosed NOV48a polypeptide (SEQ ID NO:182) encoded by SEQ ID NO:181 has 752 amino acid residues and is presented in Table 48B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV48a has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. Alternatively, NOV48a may also localize to the microbody (peroxisome) with a certainty of 0.1265, to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 48B. Encoded NOV48a protein sequence (SEQ ID NO:182).

<p>MQEGMPMRKVRSKSWKKLRYFRLQNDGMTVWHARQARQSAKPSVSI SDVETIRNGHDSSELLRSLAEELPLEQ GFTTIVFHGRRSNLDLMANSVEEAQIWMRGLQLLVLTSMQERLDQYLSDFQGRGDKNQDGKMSFQEVQR LLHLMNVEMDQEYAFSLLQAADTSQSGTLEGEFEFVQFYKALTAKRAEVQELFESFSADGQKLTLLFDFLQ EQKERDCTSELAL ELIDRYEPSDSGKLRHVLSMDGFLSYLCSKDGDI FNPACLP IYQDMTQPLNHYFICSSH NTYLVGDQLCGQSSVEGYIRALKRGCRCEVDVWDGPGSGEPVYVHGHTLT SRILFKDVVATVAQYAFQTSY PVILSLETHCSWEQQOTMARHLTEILGEQLLSTTLDGVLPTQLPSPEELRRKILVKGKLTLEEDLEYEEEE AEPELEESSELALESQFETEPEPQEQLNQNKKVSKQLQKSKPILCPALSSLVIYKLSVSVFRSFTHSKEHY HFYEISSFSSETKAKRLIKEAGNEFVQHNTWQLSRVYPSGLRTDSSNYNPQELWNAGCQMVAMNMQTAGLEM DICDGHFRQNGGCGYVLKPDFLRDIQSSFHPEKPI SPFKAQTLNQNQVISGQQLPKVDKTEGSI VDPVVKVQ IFGVRLDTARQETNYVENNGFNPNYWGQTLCFRVLVPELAMLR FVVMDDYDWSRNDFFIGQYTLPTWTCMQQGYR HIHLLSKDGISLRPASIFVYICIQEGLEGDES</p>
--

A search of sequence databases reveals that the NOV48a amino acid sequence has 619 of 752 amino acid residues (82%) identical to, and 675 of 752 amino acid residues (89%) similar to, the 764 amino acid residue ptnr:pir-id:S14113 protein from bovine (1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC 3.1.4.11) delta-2) (E = 0.0).

NOV48a is predicted to be expressed in at least Amygdala, Bone Marrow, Brain, Epidermis, Heart, Hypothalamus, Lung, Mammary gland/Breast, Pituitary Gland, Placenta, Retina, Skeletal Muscle, Small Intestine, Stomach. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV48b

In the present invention, the target sequence identified previously, NOV48a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most

downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV48b. This differs from the previously identified sequence (NOV48a) in having a deletion of 6 amino acids in one region and one amino acid at another region.

A disclosed NOV48b nucleic acid of 2341 nucleotides (also referred to as CG56743-02) encoding a Phospholipase C Delta-4-like protein is shown in Table 48C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 55-57 and ending with a TGA codon at nucleotides 2278-2280. The start and stop codons are shown in bold in Table 48C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 48C. NOV48b nucleotide sequence (SEQ ID NO:183).

```

TTTCTGACCTAGAGCTGACCACTGATCAGGGCTTGCTGCTGGTGCAGGAAGGCATGCCGATGCGCAAGGTG
AGGTCCAAAAGCTGGAAGAAGCTAAGATACTTCAGATTTCAGAATGACGGCATGACAGTCTGGCATGCACGG
CAGGCCAGGGGCAGTGCCAAGCCAGCTTCTCAATCTCTGATGTGGAGACAATACGTAATGGCCATGATTC
GAGTTGCTGCGTAGCCTGGCAGAGGAGCTCCCCCTGGAGCAGGGCTTCACCATTTGCTCTCCATGGCCGCCG
TCCAACCTGGACCTGATGGCCAACAGTGTTGAGGGGGCCAGATATGGATGCGAGGGCTCCAGCTGTTGGTG
GATCTTGTCACCAAGCATGGACCATCAGGAGCGCTGGACCAATGGCTGAGCGATTGGTTTCAACGTGGAGAC
AAAATCAGGATGGTAAGATGAGTTTCCAAGAAGTTCAGCGGTTATTCACCTAATGAATGGGAAATGGAC
CAAGAATATGCCTTCAGTCTTTTTTCAGGCAGCAGACAGTCCCAGTCTGGAACCTGGAAGGAGAAGAATTC
GTACAGTTCTATAAGGCATTGACTAAACGTGCTGAGGTGCAGGAACGTTTGAAGTTTTTCAGCTGATGGG
CAGAAGCTGACTCTGCTGGAATTTTTGGATTTCTCCTCCAAGAGGAGCAGAAGGAGAGAGACTGCACCTCTGAG
CTTGCTCTGGAACCTATTGACCGCTATGAACCTTCAGACAGTGGCAAACCTGCGGCATGTGCTGAGTATGGAT
GGCTTCTCAGCTACCTCTGCTCTAAGGATGGAGACATCTTCAACCCAGCCTGCCTCCCCATCTATCAGGAT
ATGACTCAACCCCTGAACCACTACTTCATCTGCTCTTCTCATAACACCTACCTAGTGGGGGACCAGCTTTGT
GGCCAGAGCAGCGTCGAGGGATATATACGGGCCCTGAAGCGGGGTGCCGCTGCGTGGAGGTGGATGTATGG

```

GATGGACCTAGCGGGGAACCTGTCGTTTACCACGGACACACCCCGACCTCCCGCATCCTGTTCAAAGATGTC
 GTGGCCACAGTAGCACAGTATGCCCTTCCAGACATCAGACTACCCAGTCATCTTGTCCCTGGAGACCCACTGC
 AGCTGGGAGCAGCAGCAGACCATGGCCCGTCATCTGACTGAGATCCTGGGGGAGCAGCTGCTGAGCACCAACC
 TTGGATGGGGTGCTGCCCACTCAGCTGCCCTCGCCTGAGGAGCTTCGGAGGAAGATCCTGGTGAAGGGGAAG
 AAGTTAACACTTGAGGAAGACCTGGAATATGAGGAAGAGGAAGCAGAACCTGAGTTGGAAGAGTCAGAATTG
 GCGCTGGAGTCCCAGTTGAGACTGAGCCTGAGCCCCAGGAGCAGAACCTTCAGAATAAGGACAAAAGAAG
 AAATCCAAGCCCATCTTGTGTCCAGCCCTCTCTTCCCTGGTTATCTACTTGAAGTCTGTCTCATTCCGAGC
 TTCACACATTCAAAGGAGCACTACCACTTCTACGAGATATCATCTTTCTCTGAACCAAGGCCAAGCGCCTC
 ATCAAGGAGGCTGGCAATGAGTTTGTGCAGCACAATACTTGGCAGTTAAGCCGTGTGTATCCCAGCGGCCTG
 AGGACAGACTCTTCCAACATAACCCCCAGGAACCTGGAATGCAGGCTGCCGATGGTGGCCATGAATATG
 CAGACTGCAGGGCTTGAATGGACATCTGTGATGGGCATTTCGCCCAGAATGGCGGCTGTGGCTATGTGCTG
 AAGCCAGACTTCCTGCGTGATATCCAGAGTTCTTCCACCCTGAGAAGCCCATCAGCCCTTCAAAGCCAG
 ACTCTCTTAATCCAGGTGATCAGCGGTGAGCAACTCCCCAAAGTGGACAAGACCAAGAGGGGTCCATTGTG
 GATCCATGGTGAAAGTGCAGATCTTGGCGTTGCTCTAGACACAGCACGGCAGGAGACCAACTATGTGGAG
 AACATGGTTTAAATCCATACTGGGGGCAGACACTATGTTTCCGGGTGCTGGTGCCTGAACCTGCCATGCTG
 CGTTTTGTGGTAATGGATTATGACTGGAAATCCCGAAATGACTTTATGGTTCAGTACACCCCTGCCTTGGACC
 TGCATGCAACAAGGTTACCGCCACATTACCTGCTGTCCAAAGATGGCATCAGCCTCCGCCAGCTTCCATC
 TTTGTGTATATCTGCATCCAGGAAGGCTGGAGGGGGTGAGTCCTGAGGTGGGCATTTACGCGGAAGGGTT
 GGTGTGCTGGCTTTAGACGGGGAGAAACATCTGGAAG

In a search of public sequence databases, the NOV48b nucleic acid sequence, located
 on chromosome 2, has 1069 of 1075 bases (99%) identical to a gb:GENBANK-
 ID:AK023083|acc:AK023083.1 mRNA from *Homo sapiens* (cDNA FLJ13021 fis, clone
 5 NT2RP3000742, weakly similar to 1-Phosphatidylinositol-4,5-Bisphosphate
 Phosphodiesterase Delta 1 (EC 3.1.4.11)) (E = 0.0).

The disclosed NOV48b polypeptide (SEQ ID NO:184) encoded by SEQ ID NO:183
 has 741 amino acid residues and is presented in Table 48D using the one-letter amino acid
 code. Signal P, Psort and/or Hydropathy results predict that NOV48b has no signal peptide
 10 and is likely to be localized to the mitochondrial matrix space with a certainty of 0.6523.
 Alternatively, NOV48b may also localize to the mitochondrial inner membrane with a
 certainty of 0.3462, to the mitochondrial intermembrane space with a certainty of 0.3462, or to
 the mitochondrial outer membrane with a certainty of 0.3462.

Table 48D. Encoded NOV48b protein sequence (SEQ ID NO:184).	
MPMRKVRSKSWKKLRYFRFQNDGMTVWHARQARGSAKPSFISDVETIRNGHDSSELLRSLAEELPLEQGFTI	
VFHGRRSNLDMANSVEGAQIWMRGLQLLVDLVTSMDHQERLDQWLSDFWQRGDKNQDGKMSFQEVQRLHL	
MNVEMDQEYAFSLFQAADTSQSGTLEGEFEFVQFYKALTKRAEVQELFESFSADGQKLTLEFLDFLQEEQKE	
RDCTSELALELIDRYEPSDSGKLRLHVLSDMGFLSYLCSKDGDI FNPACLP IYQDMTQPLNHYFICSSHNTYL	
VGDQLCGQSSVEGYIRALKRGCRCVEVDVWDGPSGEPVVYHGHTPTSRILFKDVVATVAQYAFQTS DYPVIL	
SLETHCSWEQQQTMARHLTEILGEQLLSTTL DGVLP TQLPSPEELRRKILVKGKKLTLEEDLEYEEEEAEPE	
LEESLALLESQFETEPEPEQNLQNKDKKKSKPILCPALSSSLVIYLSVSVFRSFTHSKEHYHFYEISSFSE	
TKAKRLIKEAGNEFVQHNTWQLSRVYPSGLRTDSSNYPQELWNAGCRMVAMNMQTAGLEMDICDGHFRQNG	
GCGYVLKPDFLRDIQSSFHPEKPI SPFKAQTL LIQVISGQQLPKVDKTEKGSIVDPLVKVQIFGVRLDTARQ	
ETNYVENNGFNPNYWGQTL CFRVLVPELAMLR FVVMDDYDWKSRNDFIGQYTL PWT CMQQGYRHIHLLSKDGIS	
LRPASIFVYICIQEGLEGGES	

15 A search of sequence databases reveals that the NOV48b amino acid sequence has 736
 of 741 amino acid residues (99%) identical to, and 737 of 741 amino acid residues (99%)

similar to, the 762 amino acid residue ptnr:TREMBLNEW-ACC:AAH06355 protein from *Homo sapiens* (Human) (Unknown (Protein For MGC:12837)) (E = 0.0).

NOV48b is predicted to be expressed in at least Heart, Stomach, Small Intestine, Bone Marrow, Skeletal Muscle, Brain, Hypothalamus, Pituitary Gland. The sequence is predicted to have the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK023083|acc:AK023083.1) a closely related *Homo sapiens* cDNA FLJ13021 fis, clone NT2RP3000742, weakly similar to 1-Phosphatidylinositol-4,5-Bisphosphate Phosphodiesterase Delta 1 (EC 3.1.4.11) homolog.

NOV48a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 48E.

Table 48E. BLAST results for NOV48a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 14249340 ref NP_116115.1 (NM_032726)	hypothetical protein MGC12837 [<i>Homo sapiens</i>]	762	736/752 (97%)	738/752 (97%)	0.0
gi 108854 pir S14113	1- phosphatidylinosi tol-4,5- bisphosphate phosphodiesterase (EC3.1.4.11) delta-2 - bovine	764	613/757 (80%)	671/757 (87%)	0.0
gi 18093100 ref NP_542419.1 (NM_080688)	phospholipase C, delta 4 [<i>Rattus norvegicus</i>]	772	550/755 (72%)	631/755 (82%)	0.0
gi 1304189 dbj BAA09046.1 (D50455)	phodpholipase C delta4 [<i>Rattus norvegicus</i>]	771	548/756 (72%)	629/756 (82%)	0.0
gi 12855950 dbj BAB30513.1 (AK016945)	data source:MGD, source key:MGI:107469, ev idence:ISS-phosph olipase C, delta 4-putative [<i>Mus musculus</i>]	447	335/430 (77%)	375/430 (86%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 48F. In the ClustalW alignment of the NOV48 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

```

1) Novel NOV48a (SEQ ID NO:182)
2) Novel NOV48b (SEQ ID NO:184)
5 3) gi|14249340|ref|NP_116115.1| (NM_032726) hypothetical protein MGC12837 [Homo sapiens] (SEQ ID NO:542)
4) gi|108854|pir||S14113 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC3.1.4.11) delta-2 - bovine (SEQ ID NO:543)
10 5) gi|18093100|ref|NP_542419.1| (NM_080688) phospholipase C, delta 4 [Rattus norvegicus] (SEQ ID NO:544)
6) gi|1304189|dbj|BAA09046.1| (D50455) phodpholipase C delta4 [Rattus norvegicus] (SEQ ID NO:545)
7) gi|12855950|dbj|BAB30513.1| (AK016945) data source:MGD, source key:MGI:107469,evidence:ISS-phospholipase C, delta 4~putative [Mus musculus] (SEQ ID NO:546)
15

```

NOV48a 1MDECMPMRKVRSKSWKKLRYFRLQDGMGVVHARQARGSAPKS 43
 NOV48b 1MPMRKVRSKSWKKLRYFRLQDGMGVVHARQARGSAPKS 39
 gi|14249340| 1 MASLLDQDLTTDODLLLMQEGTMMRKVRSKSWKKLRYFRLQDGMGVVHARQARGSAPKS 60
 gi|108854| 1 MAYLLQGRLPINODLLLMQKGTMMRKVRSKSWKKLRYFRLQDGMGVVHARQAGGRAKP 60
 gi|18093100| 1 MASQIQKLLTTNODLLLMQKGTMMRKVRIKSWKKLRYFRLQDGMGVVHGRHLESISKPT 60
 gi|1304189| 1 MASQIQKLLTTNODLLLMQKGTMMRKVRIKSWKKLRYFRLQDGMGVVHGRHLESISKPT 60
 gi|12855950| 1 MTSQIQDLLATDODLLLMQEGTMMRKVRIKSWKKLRYFRLQDGMGVVHGSOPESMPKPT 60

			70	80	90	100	110	120																																																							
																																																									
30	NOV48a	44	V	S	I	S	D	V	E	T	I	R	N	G	H	D	S	E	L	L	R	S	L	A	E	E	L	P	L	E	Q	G	F	T	I	V	F	H	G	R	R	S	N	L	D	M	A	N	S	V	E	E	A	Q	I	W	M	R	G	L	Q		103
	NOV48b	40	F	S	I	S	D	V	E	T	I	R	N	G	H	D	S	E	L	L	R	S	L	A	E	E	L	P	L	E	Q	G	F	T	I	V	F	H	G	R	R	S	N	L	D	M	A	N	S	V	E	G	A	Q	I	W	M	R	G	L	Q		99
	gi 14249340 	61	F	S	I	S	D	V	E	T	I	R	N	G	H	D	S	E	L	L	R	S	L	A	E	E	L	P	L	E	Q	G	F	T	I	V	F	H	G	R	R	S	N	L	D	M	A	N	S	V	E	E	A	Q	I	W	M	R	G	L	Q		120
	gi 108854 	61	F	S	I	S	D	V	T	V	R	C	H	E	S	E	L	L	R	N	L	A	E	E	F	P	L	E	Q	G	F	T	I	V	F	H	G	R	R	S	N	L	D	L	V	A	N	S	V	Q	E	A	Q	I	W	M	Q	G	L	Q		120	
	gi 18093100 	61	F	S	I	S	D	V	E	R	I	R	K	G	D	S	E	L	L	R	L	V	E	E	F	P	L	E	Q	G	F	T	I	V	F	N	G	R	R	P	N	L	D	L	V	A	N	S	V	E	E	A	Q	I	W	M	R	G	L	Q		120	
	gi 1304189 	61	F	S	I	S	D	V	E	R	I	R	K	G	D	S	E	L	L	R	L	V	E	E	F	P	L	E	Q	G	F	T	I	V	F	N	G	R	R	P	N	L	D	L	V	A	N	S	V	E	E	A	Q	I	W	M	R	G	L	Q		120	
35	gi 12855950 	61	F	S	I	S	D	V	E	R	I	R	K	G	D	S	E	L	L	R	L	V	E	E	F	P	L	E	Q	G	F	T	I	V	F	H	G	R	R	P	N	L	D	L	V	A	N	S	V	E	E	A	Q	I	W	M	R	G	L	Q		120	

130 140 150 160 170 180
 40 NOV48a 104 LLVDLVTSMDHQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 163
 NOV48b 100 LLVDLVTSMDHQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 159
 gi | 14249340 | 121 LLVDLVTSMDHQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 180
 gi | 108854 | 121 LLVGFEVTNMDQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 180
 gi | 18093100 | 121 LLVDLVARMNYQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 180
 gi | 1304189 | 121 LLVDLVARMNYQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 180
 45 gi | 12855950 | 121 LLVDLVASMDHQERLDQYLSDFWFGDKNQDGKMSFQEVQRLLHLMNVEMDQEYAFSLFQ 180

Accession	Position	Sequence	Position	Sequence
NOV48a	164	AADTSQSGTLEGEFVQFYKALTKRAEVOELFESFSDGQKLTLLLEFDLFLQEEQKERDC	223	
NOV48b	160	AADTSQSGTLEGEFVQFYKALTKRAEVOELFESFSDGQKLTLLLEFDLFLQEEQKERDC	219	
gi 14249340	181	AADTSQSGTLEGEFVQFYKALTKRAEVOELFESFSDGQKLTLLLEFDLFLQEEQKERDC	240	
gi 108854	181	TADTSQSGTLEGEFVFEFYKSLTORPEVOELFEKFSDDGQKLTLLLEFVDFLQEEQKEGER	240	
gi 18093100	181	EADVVSQSNITLDSEEFVQFYKALTKRTEIEELFENFSSDKOKLTLLLEFVDFLREEQKESDH	240	
gi 1304189	181	EADVVSQSNITLDSEEFVQFYKALTKRTEIEELFENFSSDKOKLTLLLEFVDFLREEQKESDH	240	
gi 12855950	181	EADVVSQSNITLDSEEFVQFYKALTKRTEIEELFENFSSDKOKLTLLLEFVDFLREEQKESDH	240	

		250	260	270	280	290	300	
60	NOV48a	224	TSELALELIDRYEPSDS	260	SKLRHVLSMDGFLSYLCSKDGD	280	IFNPACLPPIYQDMTQPLNHYF	283
	NOV48b	220	TSELALELIDRYEPSDS	260	SKLRHVLSMDGFLSYLCSKDGD	280	IFNPACLPPIYQDMTQPLNHYF	279
	gi 14249340	241	TSELALELIDRYEPSDS	260	SKLRHVLSMDGFLSYLCSKDGD	280	IFNPACLPPIYQDMTQPLNHYF	300
	gi 108854	241	ASDLALELIDRYEPSDS	260	SKLRHVLSMDGFLSYLCSKDGD	280	IFNPACLPPIYQDMTQPLNHYF	300
	gi 18093100	241	SSDLALKLIDRYEPSSE	260	SKLRHVLSMDGFLSYLCSADGNI	280	FNPDCLPPIYQDMTQPLSHYF	300
65	gi 1304189	241	SSDLALKLIDRYEPSSE	260	SKLRHVLSMDGFLSYLCSADGNI	280	FNPDCLPPIYQDMTQPLSHYF	300
	gi 12855950	241	APDLALELIDRYEPSSE	260	SKLRHVLSMDGFLSYLCSKDGD	280	IFNPACLPPIYQDMTQPLSHYF	300

421

730 740 750 760 770

NOV48a 694 MDYDWKSRNDFIGQYTLPWTCMQGYRHHLLSKDGISLRPASIFVYICIQEGLEGDES 752

NOV48b 683 MDYDWKSRNDFIGQYTLPWTCMQGYRHHLLSKDGISLRPASIFVYICIQEGLEGGES 741

5 gi|14249340| 704 MDYDWKSRNDFIGQYTLPWTCMQGYRHHLLSKDGISLRPASIFVYICIQEGLEGDES 762

gi|108854| 706 KDYDWKSRNDFIGQYTLPWTCMQGYRHHLLSKDGLSLHPASIFVYICTQEVSEEAES 764

gi|18093100| 714 KDYSRTSRNNFIGQYTLPWTCMKHGYRHVSLLSKDGTSLHPASIFVYTCMQEDLDMDEP 772

gi|1304189| 713 KDYSRTSRNNFIGQYTLPWTCMKHGYRHVSLLSKDGTSLHPASIFVYTCMQEDLDMDEP 771

10 gi|12855950| 447 ----- 447

Tables 48G-N lists the domain descriptions from DOMAIN analysis results against NOV48. This indicates that the NOV48 sequence has properties similar to those of other proteins known to contain this domain.

Table 48G Domain Analysis of NOV48

gnl|Smart|smart00148, PLCXc, Phospholipase C, catalytic domain (part); domain X; Phosphoinositide-specific phospholipases C. These enzymes contain 2 regions (X and Y) which together form a TIM barrel-like structure containing the active site residues. Phospholipase C enzymes (PI-PLC) act as signal transducers that generate two second messengers, inositol-1,4,5-trisphosphate and diacylglycerol. The bacterial enzyme appears to be a homologue of the mammalian PLCs. (SEQ ID NO:840)

CD-Length = 145 residues, 100.0% aligned

Score = 202 bits (514), Expect = 6e-53

NOV48: 273 QDMTQPLNHYFICSSHNTYLVDGQLCGQSSVEGYIRALKRGCRCEVDVWDGPGSEPVVY 332

20 Sbjct: 1 QDMSKPLSHYFINSSHNTYLTGKQLWGESSVEGYIQALKHGCRCVELDCWDGPDGEPVIY 60

NOV48: 333 HGHTLTSRILFKDVVATVAQYAFQTSYPVILSLETHCSWEQQQTMARHLTEILGEQLLS 392

25 Sbjct: 61 HGHTFTLPIKLSEVLEAIKKFAFVTSYPVILSLENHCSPDQQAQMAQMFKEIFGDLLYT 120

NOV48: 393 TTLDGVLPTQLPSPEELRRKILVKGK 418

Sbjct: 121 PPTTSSL-EYLPSPQLKGKILLKGK 145

Table 48H Domain Analysis of NOV48

gnl|Pfam|pfam00388, PI-PLC-X, Phosphatidylinositol-specific phospholipase C, X domain. This associates with pfam00387 to form a single structural unit. (SEQ ID NO:841)

CD-Length = 145 residues, 100.0% aligned

Score = 192 bits (489), Expect = 4e-50

NOV48: 274 DMTQPLNHYFICSSHNTYLVDGQLCGQSSVEGYIRALKRGCRCEVDVWDGPGSEPVVYH 333

30 Sbjct: 1 DMSIPLSHYFISSSHNTYLTGKQLWGKSQVESYRQQLDHGCRCELDWDGPDDEPIIYH 60

NOV48: 334 GHTLTSRILFKDVVATVAQYAFQTSYPVILSLETHCSWEQQQTMARHLTEILGEQLLST 393

35 Sbjct: 61 GGTFTLEIKLDVLEAIKDFLKTSPYPIILSLENHCNSDQQRKMAKYFEEIFGDYLLTK 120

NOV48: 394 TLDGVLPTQLPSPEELRRKILVKGKK 419

Table 48L Domain Analysis of NOV48

gnl|Smart|smart00239, C2, Protein kinase C conserved region 2 (CalB); Ca2+-binding motif present in phospholipases, protein kinases C, and synaptotagmins (among others). Some do not appear to contain Ca2+-binding sites. Particular C2s appear to bind phospholipids, inositol polyphosphates, and intracellular proteins. Unusual occurrence in perforin. Synaptotagmin and PLC C2s are permuted in sequence with respect to N- and C-terminal beta strands. SMART detects C2 domains using one or both of two profiles. (SEQ ID NO:845)
CD-Length = 101 residues, 100.0% aligned
Score = 83.6 bits (205), Expect = 4e-17

NOV48: 618 TLLNQVISGQQLPKVDKTEGSIVDPLVKVQIFGVRLDTARQETNYVENNGFNPNYWGQTL 677
|| ++|| + || || || ||| + | + + | |+ ||| +|
Sbjct: 1 TLTVKIISARNLPPKDKG---GKSDPYVKVSLDGDPREKKK--TKVVKNL-NPVWNETF 54
NOV48: 678 CFRVLVPELAMLRFFVMDYDWKSRNDFIGQYTLPTWTCMQQGYRHIHL 724
| | |||+ | | | ||+|||+ |+| + + | || |
Sbjct: 55 EFEVPPPELSELEIEVYDKDRFSRDDFIGRVTIPLSDLLGGRHEKL 101

Table 48M Domain Analysis of NOV48

gnl|Smart|smart00233, PH, Pleckstrin homology domain.; Domain commonly found in eukaryotic signalling proteins. The domain family possesses multiple functions including the abilities to bind inositol phosphates, and various proteins. PH domains have been found to possess inserted domains (such as in PLC gamma, syntrophins) and to be inserted within other domains. Mutations in Brutons tyrosine kinase (Btk) within its PH domain cause X-linked agammaglobulinaemia (XLA) in patients. Point mutations cluster into the positively charged end of the molecule around the predicted binding site for phosphatidylinositol lipids. (SEQ ID NO:846)
CD-Length = 104 residues, 87.5% aligned
Score = 47.8 bits (112), Expect = 2e-06

NOV48: 9 KVRSKSWKKLRYFRLQNDGMTVWHARQARGSAKPSVSIS-DVETIRNGHDSSELLRSLAE 67
||||| ||| | | + + +++ + |+|| || |+| ||+
Sbjct: 12 SGGKKSWKK-RYFVLFNGVLLYYKSKKKSSSKPKGSIPLSGCTVREAPDS-----S 63
NOV48: 68 LPLEQGFTIVFHGRRSNLDLMANSVEEAQIWMRGLQLLVD 107
+ | || |++ || | || + |+ | + +
Sbjct: 64 DKKKNCFEIVTPDRKT-LLLQAESEERKEWVEALRKAIA 102

Table 48N Domain Analysis of NOV48

gnl|Pfam|pfam00169, PH, PH domain. PH stands for pleckstrin homology (SEQ ID NO:847)
CD-Length = 100 residues, 88.0% aligned
Score = 45.1 bits (105), Expect = 1e-05

NOV48: 9 KVRSKSWKKLRYFRLQNDGMTVWHARQARGSAKPSVSI DVETIRNGHDSSELLRSLAEEL 68
|+ | ||| ||| ||| + + ++ | |+ +| + |
Sbjct: 12 TVKKKRWKK-RYFFLFNDVLIYYKDKKKSYPKGSIPLSGCS-----VEDVPDSEF 61
NOV48: 69 PLEQGFTIVFHGRRSNLDLMANSVEEAQIWMRGLQLLV 106
| + + | || | ||| ||| ++ +| +
Sbjct: 62 KRPNCFQLRSRDGKETFILQAESEERQDWIKAIQSAI 99

25

30

include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV48 protein of the invention includes the Phospholipase C Delta-4-like protein whose sequence is provided in Table 48B or 48D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 48B or 48D while still encoding a protein that maintains its Phospholipase C Delta-4-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 28 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Phospholipase C Delta-4-like protein (NOV48) is a member of a "Phospholipase C Delta-4 family". Therefore, the NOV48 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV48 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Cardiomyopathy, Atherosclerosis, Hypertension, Congenital heart defects, Aortic stenosis, Atrial septal defect (ASD), Atrioventricular (A-V) canal defect, Ductus arteriosus, Pulmonary stenosis, Subaortic stenosis, Ventricular septal defect (VSD), valve diseases, Tuberous sclerosis, Scleroderma, Obesity, Transplantation, Osteoporosis, Hypercalcaemia, Arthritis, Ankylosing spondylitis, Scoliosis, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, and/or other diseases and pathologies.

NOV48 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV48 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV48 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV49

A disclosed NOV49 nucleic acid of 1588 nucleotides (also referred to as CG56739-01) encoding a Leukotriene-B4 Omega- Hydroxylase -like protein is shown in Table 49A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 2-4 and ending with a TGA codon at nucleotides 1577-1579. The start and stop codons are shown in bold in Table 49A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 49A. NOV49 nucleotide sequence (SEQ ID NO:185).

GATGTCGCTGCTGAGCCTGTCCTGGCTGGGCCCTCGGGCCGGTGGCAGCATCCCCATGGCTGCTCCTGCTGCT
GGTCGGGGCCCTCCTGGCTCCTGGGCCCGTGTCTGGCCCTGGACGTACGCCTTCTATGACAACATGCCACCGCCT
CCAGTGTTCAGCAGCCTCCAAACGGAACCTGCTTTTAGGTACCTGAGCCTGGTGCGGGGCAATGAGGA
GGACATGAGGCTGATGGAGGATCTGGGCCACTACTCCGTGATGTCCAACCTCTGGTGGCTTGGGTCTTTCTA
CCCTGTCCTGCATCTCGTCCACCCTACGTTCACTGCCCCCTGTGCTCCAGGCTTCAGCTGCTGTTGCACTCAA
GGATATGAGTTTCTATGGCTTCCTGAAGCCCTGGCTGGGTCTCTGATGGGCTCCTGATTAGTGCCGGTGACAA
GTGGAGATGGCACCGCCACCTGCTCACACCTGCCTTCCACTTCAAAATCCTGAAGCCCTATGTGAAGATTTT
CAATGAGAGCACGAACATCATGCACGCCAAATGGCAACGCCTGGCCCTGGAGGGCAGTGTCCGTCTGGAAAT
GTTTGAGCACATCAGCCTCATGACCTTGGACAGTCTGCAGAAATGCATCTTCAGCTTTGACAGCAATTGTCA
GGAGAAGCCCAGCGAATATATTGATGCCATCTTGGAGCTCAGTGCCCTCAGTCTGAAACGGCACCAGCACAT
CTTCCTGCTCACGGACTTCTGTACTTCCTCACTCCCAATGGGCGACGCTTCTGCAGGGCCTGTGACATAGT
GCACAACTTCACAGATGCTGTATCCAGGAGCGGCGTCGACCCCTCACTAGCCAGGGTGTCTGATGACTTCCT
GCAGGCCAAGGCCAAGTCCAAGACTTTGGACTTCATTGACGTGCTCTTGCTGGCCAAGGATGAAAATGGAAA
GAAGTTGTGATGAGAACATAAGAGCGGAGGCTGACACCTTCATGTCTGGGGCCATGACACCTCGGCCAG
TGGTCTCTCCTGGGTCTGTACAACCTCGCGAGGTACCCAGAATACCAGGAGCACTGCCGACAGGAGGTGCA
AGAGCTCCTGAAGAACGGTGATCCTAAAGAGATTGAATGGGATGACCTGGCCAGTTGCCCTTCTGACCAT
GTGCCTGAAGGAGAGCCTGCGGCTGCATTCCCCAGTCTCCAGGATCCACCGCTGTGCCCCCAGGACGGGGT
GCTCCCGGATGGCCGGGTGATCCCCAAAGGTAACACTTGCACCATCAGCATCTTTGGGATCCATCACAACCC
TTCAGTCTGGCCGGACCCGAGGTGATGACCCCTTTCGCTTCGACCCAGAAAATCTCCAGAAGACATCACC
TCTGGCTTTTATTCCTTCTCAGCAGTGCCAGGAACTGCATCGGCCAGACGTTCCGCATGGCTGAGATGAA
GGTGGTCTGGCGCTCACGCTGCTGCGCTTCCGCGTCTGCGGACCCAGCGGAGCCCCGAGGAAGCTGGA
GCTGATCGTGCGCGGAGGATGGACTTTGGCTACGGGTGGAGCCCCTGAGCGCGGATCTGCAGTGACCCAC
CACT

In a search of public sequence databases, the NOV49 nucleic acid sequence, located on chromosome 19, has 1320 of 1584 bases (83%) identical to a gb:GENBANK-

ID:HUMLB4OH|acc:D26480.1 mRNA from *Homo sapiens* (Human mRNA for leukotriene B4 omega-hydroxylase, complete cds) (E = 9.7e⁻²³⁷).

The disclosed NOV49 polypeptide (SEQ ID NO:186) encoded by SEQ ID NO:185 has 525 amino acid residues and is presented in Table 49B using the one-letter amino acid code.

Signal P, Psort and/or Hydropathy results predict that NOV49 has a signal peptide and is likely to be localized extracellularly with a certainty of 0.8200. Alternatively, NOV49 may also localize to the lysosome (lumen) with a certainty of 0.4520, to the microbody (peroxisome) with a certainty of 0.1611, or to the endoplasmic reticulum (membrane) with a certainty of 0.1000. The most likely cleavage sit for NOV49 is between positions 36 and 37: VLA-WT.

Table 49B. Encoded NOV49 protein sequence (SEQ ID NO:186).

MSLLSLSWLGLGPVAASPWLLLLLVGASWLLARVLAWTYAFYDNCHRLQCFFQPPKRNCFGLGHLVLVRGNEE
DMRLMEDLGHYFRDVQLWWLGSFYFVPLHLVHPTFTAPVLQASAAVALKDMSFYGFLKPWLGPDGLLIISAGDK
WRWHRHLLTPAFHFKILKPYVKIFNESTNIMHAKWQRLALEGSVRLEMFEHISLMTLDSLQKCIFSDSNQC
EKPSEYIDAILLSALSILKRHOHI FLLTDFLYFLT PNGRRFCRACDIVHNFTDAVIQERRRITLSQGVDDFL
QAKAKSKTLDFIDVLLAKDENGKKLSDENIRAEADTFMSGGHDTSASGLSWVLYNLARYPEYQEHCRQEVQ
ELLKNGDPKEIEWDDLAQLPFLTMCLKESLRLHSPVSRIHRCCPDGVLDPGRVIPKGNCTCTISIFGIHNHP
SVWPDPEVYDPRFDPENLQKTSPLAFIPFSAVPRNCIGQTFAMAEMKVVLALTLLRFRVLPDHAEP RRKLE
LIVRAEDGLWLRVEPLSADLQ

A search of sequence databases reveals that the NOV49 amino acid sequence has 397 of 521 amino acid residues (76%) identical to, and 444 of 521 amino acid residues (85%) similar to, the 520 amino acid residue ptnr:SWISSPROT-ACC:Q08477 protein from *Homo sapiens* (Human) (CYTOCHROME P450 4F3 (EC 1.14.13.30) (CYP11B) (Leukotriene-B4 Omega- Hydroxylase) (Leukotriene-B4 20-Monooxygenase) (Cytochrome P450- LTB-Omega)) (E = $4.3e^{-219}$).

NOV49 is predicted to be expressed in at least Prostate. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in Bone Marrow, Peripheral Blood, Brain, Colon, Coronary Artery, Hippocampus, Kidney, Kidney Cortex, Liver, Lymph node, Pituitary Gland, and Prostate because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HUMLB4OH|acc:D26480.1) a closely related Human mRNA for leukotriene B4 omega-hydroxylase, complete cds homolog.

NOV49 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 49C.

Table 49C. BLAST results for NOV49

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17456512 ref XP_065069.1 (XM_065069)	similar to CYTOCHROME P450 4F6 (CYPIVF6) (H. sapiens) [<i>Homo sapiens</i>]	491	452/495 (91%)	455/495 (91%)	0.0
gi 14767705 ref XP_029072.1 (XM_029072)	cytochrome P450, subfamily IVF, polypeptide 3 [<i>Homo sapiens</i>]	520	399/521 (76%)	446/521 (85%)	0.0
gi 4503241 ref NP_00887.1 (NM_000896)	cytochrome P450, subfamily IVF, polypeptide 3;leukotriene B4 omega hydroxylase; leukotriene-B4 20-monooxygenase; cytochrome P450- LTB-omega [<i>Homo sapiens</i>]	520	397/521 (76%)	444/521 (85%)	0.0
gi 2997737 gb AAC08589.1 (AF054821)	cytochrome P-450 [<i>Homo sapiens</i>]	520	395/521 (75%)	440/521 (83%)	0.0
gi 1706095 sp P51871 CPF6_RAT	CYTOCHROME P450 4F6 (CYPIVF6)	537	392/521 (75%),	443/521 (84%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 49D. In the ClustalW alignment of the NOV49 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 49D. ClustalW Analysis of NOV49

- 1) Novel NOV49 (SEQ ID NO:186)
- 2) gi|17456512|ref|XP_065069.1| (XM_065069) similar to CYTOCHROME P450 4F6 (CYPIVF6) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:547)
- 3) gi|14767705|ref|XP_029072.1| (XM_029072) cytochrome P450, subfamily IVF, polypeptide 3 [*Homo sapiens*] (SEQ ID NO:548)
- 4) gi|4503241|ref|NP_00887.1| (NM_000896) cytochrome P450, subfamily IVF, polypeptide 3;leukotriene B4 omega hydroxylase; leukotriene-B4 20-monooxygenase; cytochrome P450-LTB-omega [*Homo sapiens*] (SEQ ID NO:549)
- 5) gi|2997737|gb|AAC08589.1| (AF054821) cytochrome P-450 [*Homo sapiens*] (SEQ ID NO:550)
- 6) gi|1706095|sp|P51871|CPF6_RAT CYTOCHROME P450 4F6 (CYPIVF6) (SEQ ID NO:551)

			10	20	30	40	50	60
NOV49	1	MSLLSLSWLG	LPV	MAASPWLLLLLVGASWLLAR	VLAWTYAFYDNC	HLRC	CFQPPKRN	CF
gi 17456512	1	MSLLSLSWLG	LPV	MAASPWLLLLLVGASWLLAR	VLAWTYAFYDNC	HLRC	CFQPPKRN	CF
gi 14767705	1	MPQLSLSWLG	LPV	MAASPWLLLLLVGASWLLAR	ILAWTYTFYDNC	CLRC	CFQPPKRN	WF

gi 4503241	1	MPQLSLSSGLWPMMAASPWLLLLLVGASWLLARILAWTYTFYDNCCRLRCFPQPPKRNWF	60
gi 2997737	1	MPQLSLSSGLWPMMAASPWLLLLLVGASWLLARILAWTYTFYDNCCRLRCFPQPPKRNWF	60
gi 1706095	1	MLQLSLSRLCMGSLTASPWLLLLLVGASWLLARILAWTYTFYDNCCRLRCFPQPPKPSWF	60

5		70	80	90	100	110	120	
NOV49	61	LGHLSLVRGNEEDMRLMEDLGHYFRDVLWNLGSFYFVHLVHPTFTAPVLQASAAVALK	120					
gi 17456512	1	-----MRCNEEDMRLMEDLGHYFRDVLWNLGSFYFVHLVHPTFTAPVLQASAAVALK	54					
gi 14767705	61	LGHLCGLIHSSEEGLLYTQSLACTFGDMCCWVVGFWHAIVRIFFHPTYIKPVLFAFAAIVPK	120					
gi 4503241	61	LGHLCGLIHSSEEGLLYTQSLACTFGDMCCWVVGFWHAIVRIFFHPTYIKPVLFAFAAIVPK	120					
gi 2997737	61	LGHLCGLVTPTECGMRVLTQLVATYVPGFKVVMGPIFPVIRFCHPNIRSVINASAAIVPK	120					
gi 1706095	61	WGHLTLMKNNBEGMQFIAHLGRNFRDHLSSWVGFPVYPIRLVHPNVIAELLQASAAVAPK	120					

15		130	140	150	160	170	180	
NOV49	121	DMSFYGFLKPWLGPDGLLSAGDKWRWRRHLLTPAFHFKILKPYVKIFNESVNIMHAKWQ	180					
gi 17456512	55	DMSFYGFLKPWLGPDGLLSAGDKWRWRRHLLTPAFHFKILKPYVKIFNESVNIMHAKWQ	113					
gi 14767705	121	DKVFYSFLKPWLGPDGLLSAGDKWRWRRHLLTPAFHFKILKPYVKIFNESVNIMHAKWQ	179					
gi 4503241	121	DKVFYSFLKPWLGPDGLLSAGDKWRWRRHLLTPAFHFKILKPYVKIFNESVNIMHAKWQ	179					
gi 2997737	121	DKVFYSFLKPWLGPDGLLSAGDKWRWRRHLLTPAFHFKILKPYVKIFNESVNIMHAKWQ	179					
gi 1706095	121	EMTLYGFLKPWLGPDGLLSAGDKWNRHRRLLTPAFHFDILKSYVKIFNKSVMIMHAKWQ	179					

25		190	200	210	220	230	240	
NOV49	181	RLALEGSVRLEMFHEHISLMTLDSLOKCIFSFDSNCOEKPSEYIDAILLSALSILKRHQHI	240					
gi 17456512	114	RLALEGSVRLEMFHEHISLMTLDSLOKCIFSFDSNCOQ---EYIDAILLSALSILKRHQHI	170					
gi 14767705	180	LLASEGSARLDMFEHISLMTLDSLOKCVFSFDSHCQEKPSYIAAILLSALVTKRHQOI	239					
gi 4503241	180	LLASEGSARLDMFEHISLMTLDSLOKCVFSFDSHCQEKPSYIAAILLSALVTKRHQOI	239					
gi 2997737	180	LLASKGYARLDMFEHISLMTLDSLOKCVFSFDSHCQEKPSYIAAILLSALVTKRHQOI	239					
gi 1706095	180	RLTAKGSARLDMFEHISLMTLDSLOKCIFSFDSNCOESNSEYIAAILLSLVKRRQRP	239					

35		250	260	270	280	290	300	
NOV49	241	FLLTDFLYELTPNGRREFCRACDIVHNFDTDAVIQERRRITLSQGVDDFLQAKAKSKTLDFI	300					
gi 17456512	171	FLLTDFLYELTPNGRREFCRACDIVHNFDTDAVIQERRRITLSQGVDDFLQAKAKSKTLDFI	230					
gi 14767705	240	LLYIDFLYELTPDGRFRACRLVHDFDTDAVIQERRRITLSQGVDDFLQAKAKSKTLDFI	299					
gi 4503241	240	LLYIDFLYELTPDGRFRACRLVHDFDTDAVIQERRRITLSQGVDDFLQAKAKSKTLDFI	299					
gi 2997737	240	LLYIDFLYELTPDGRFRACRLVHDFDTDAVIQERRRITLSQGVDDFLQAKAKSKTLDFI	299					
gi 1706095	240	FLYLDFLYELTADGRFRACDIVHNFDTDAVIRERRSTLNTQGVDFLQAKAKSKTLDFI	299					

45		310	320	330	340	350	360	
NOV49	301	DVLLLLAKDENGGKLSDENIRAEADTFMSGGHDTSASGLSWVLYNLARYPEYQEHCRQEVQ	360					
gi 17456512	231	DVLLLLAKDENGGKLSDENIRAEADTFMSGGHDTSASGLSWVLYNLARYPEYQEHCRQEVQ	290					
gi 14767705	300	DVLLLLSKDEDDGKLSDEDIRAEADTFMFEHGDHDTASGLSWVLYHLAKHPEYQERCRCQEVQ	359					
gi 4503241	300	DVLLLLSKDEDDGKLSDEDIRAEADTFMFEHGDHDTASGLSWVLYHLAKHPEYQERCRCQEVQ	359					
gi 2997737	300	DVLLLLSKDEDDGKLSDEDIRAEADTFMFEHGDHDTASGLSWVLYHLAKHPEYQERCRCQEVQ	359					
gi 1706095	300	DVLLLLAKDEHGKGLSDVDIRAEADTFMFEHGDHDTASGLSWVLYNLARYPEYQERCRCQEVQ	359					

55		370	380	390	400	410	420	
NOV49	361	ELLKNGDPKEIEWDDLAQLPFLTMCIKESLRLHSPVSRIRHCCPDQGVLPDGRVIPKGN	420					
gi 17456512	291	ELLKNGDPKEIEWDDLAQLPFLTMCIKESLRLHSPVSRIRHCCPDQGVLPDGRVIPKGN	350					
gi 14767705	360	ELLKDREPKIEWDDLAQLPFLTMCIKESLRLHPPVPAVSRCCQDQIVLPDGRVIPKGI	419					
gi 4503241	360	ELLKDREPKIEWDDLAQLPFLTMCIKESLRLHPPVPAVSRCCQDQIVLPDGRVIPKGI	419					
gi 2997737	360	ELLKDREPKIEWDDLAQLPFLTMCIKESLRLHPPVPAVSRCCQDQIVLPDGRVIPKGI	419					
gi 1706095	360	ELLKDREPKIEWDDLAQLPFLTMCIKESLRLHPPVLLTSRCCSQDQIVLPDGRVIPKGN	419					

65		430	440	450	460	470	480	
NOV49	421	CLISIFGTHHNPSVWPDPE-----VYDPFRFDPENLQKTSPLAFIPFSAGPR	467					
gi 17456512	351	CLISIFGTHHNPSVWPDPEVLPPLPPSPSRGLVYDPFRFDPENLQKTSPLAFIPFSAGPR	410					
gi 14767705	420	CLISVFGTHHNPAVWPDPE-----VYDPFRFDPKNIKERSPLAFIPFSAGPR	466					
gi 4503241	420	CLISVFGTHHNPAVWPDPE-----VYDPFRFDPKNIKERSPLAFIPFSAGPR	466					
gi 2997737	420	CLISVFGTHHNPAVWPDPE-----VYDPFRFDPKNIKERSPLAFIPFSAGPR	466					
gi 1706095	420	CVISIFGVHNNPSVWPDPE-----VYNPFRFDPENLQKRSPLAFIPFSAGPR	466					

70		490	500	510	520	530	540	
NOV49	467	-----NCIGQTFAMAEMKVVLALTLLRFRVLPDHAEPRRKLE	504					

5
gi|17456512| 411 GSRRDAGAGMVLGTVRVPTSPGNCIGQTFAMAEMKVVLALTLLRFRVLPDHAEP RRKLE 470
gi|14767705| 466 -----NCIGQAFAMAEMKVVLGLTLLRFRVLPDHT EPRRKPE 503
gi|4503241| 466 -----NCIGQAFAMAEMKVVLGLTLLRFRVLPDHT EPRRKPE 503
gi|2997737| 466 -----NCIGQAFAMAEMKVVLGLTLLRFRVLPDHT EPRRKPE 503
gi|1706095| 466 -----NCIGQTFAMSEIKVALALTLLRF CVLPDDKEPRRKPE 503

10
NOV49 505 LIVRAEDGLWLRVEPLSADLQ----- 525
gi|17456512| 471 LIVRAEDGLWLRVEPLSADLQ----- 491
gi|14767705| 504 LVLRAEGGLWLRVEPLS----- 520
gi|4503241| 504 LVLRAEGGLWLRVEPLS----- 520
gi|2997737| 504 LVLRAEGGLWLRVEPLS----- 520
gi|1706095| 504 LILRAEGGLWLRVEPLS TVTSQLPVDLLAHPPTS 537

Table 49F lists the domain description from DOMAIN analysis results against NOV49.

This indicates that the NOV49 sequence has properties similar to those of other proteins known to contain this domain.

Table 49F Domain Analysis of NOV49

gnl|Pfam|pfam00067, p450, Cytochrome P450. Cytochrome P450s are involved in the oxidative degradation of various compounds. Particularly well known for their role in the degradation of environmental toxins and mutagens. Structure is mostly alpha, and binds a heme cofactor. (SEQ ID NO:848)
CD-Length = 445 residues, 98.9% aligned
Score = 308 bits (790), Expect = 4e-85

25
NOV49: 55 PKRNCFLGHLSLVRGNEEDMRLMEDLGHYFRDVQLWWLGSFYFVLHLVHPTFTAPVLQAS 114
| | +|+| + + + +| + | +|| | | + + | ||
Sbjct: 4 PPPLPLIGNLLQLGRG--PIHSLTELRRKKYGPVFTLYLGP-RPVVVVTGPEAVKEVLIDK 60
NOV49: 115 AAVALKDMSFYGFLKPWLGPDGLLISAGDKWRWHRHLLTPAFHFKILKPYVKIFNESTNI 174
| | | | | | +| | | +|| | | | | | + | | +
Sbjct: 61 GEEFAGRGDFPVF--PWLG-YGILFSNGPRWRQLRRLTLRF-FGMGKRS-KLEERIQEE 115
30
NOV49: 175 MHAKWQRLALEGSRVLEMFEHISLMTLDSLQKCIFSFSNQCQEKPS E-YIDAILELSALS 233
+|| | + + + | ++ | + + + | + + + | | + | |
Sbjct: 116 ARDLVERLRKEQGSPIDITELLAPAPLNVICSLLEFGVRFDYEDPEFLKLIDKLNELFFLV 175
35
NOV49: 234 LKRHQHIFLLTDFLYFLTPNGRRRFCRACDIVHNFTDAVIQERRR TLTSQGVDDFLQAKAK 293
| | | | +| + | + | + + + | +|+|| | |
Sbjct: 176 SPWQQ----LLDFFRYLPGSHRKAFKAADLKDYLDKLIEERRETLE-----P 219
40
NOV49: 294 SKTLDFIDVLLL-AKDENGKKLSDENIRAEADTFMSGGHDTSASGLSWVLYNLARYPEYQ 352
| |+| |+| | | | +|+|| ++| + | |++| | | | |++|| |
Sbjct: 220 GDPRDFLDSLLEAKREGGSELTDEELKATVLDLLFAGTDTTSSTLSWALYLLAKHPEVQ 279
NOV49: 353 EHCRQEVQELLKNGDPKEIEWDDLAQLPFLTMCLKESLR LHSPV-SRIHRCCPDQGVLPD 411
| |+| ++| + + +| +|+|| ++| +|+|| | | + | + | + |
Sbjct: 280 AKLREEIDEVI--GRDRSPTYDDRANMPYLD AVIKETLRLHPVVPLLPRVATEDTEI-D 336
45
NOV49: 412 GRVIPKGNTCTISIFGIHNPVSWPDPEVYDPFRFDPENLQKTSPLAFIPFSAVPRNCIG 471
+|+|| + + + + +| +|+|| +|+|| | | + | +|+|| | | +|
Sbjct: 337 GYLIPKGT LIVINLYSLHRDPKVPNPPEFDPERFLDENGKFKKSYAFLPFGAGPRNCLG 396
50
NOV49: 472 QTFAMAEMKVVLALTLLRFRV-LPDHAEP RRKLELIVRAEDGLWLRV 517
+ | |+ + || | | + | + + + + + + +
Sbjct: 397 ERLARMELFLFLATLLQRFELVPPGDIPLTPKPLGLPSKPPLYQL 443

Leukotrienes are a group of bioactive compounds that play important roles in such processes as inflammation. Kikuta et al. (1993) (J. Biol. Chem. 268: 9376-9380) isolated a cDNA for the human leukotriene B4 omega-hydroxylase (LTB4H), an enzyme which catalyzes the omega-hydroxylation of leukotriene B4. Their cDNA encoded a 520-amino acid protein with a predicted molecular weight of 59,805 Da. The deduced amino acid sequence contains a cysteine in the conserved heme-binding domain near the C-terminus, which is a characteristic feature of the cytochrome P450 superfamily; the protein shares 31 to 44% similarity with CYP4A, CYP4B, and CYP4C. Kikuta et al. (1993) (J. Biol. Chem. 268: 9376-9380) detected transcript from the LTB4H gene in polymorphonuclear leukocytes and leukocytes. Kikuta et al. (1998) (DNA Cell Biol. 17: 221-230) determined that the CYP4F3 gene contains 13 exons and spans approximately 22.2 kb. By fluorescence in situ hybridization, they mapped the CYP4F3 gene to 19p13.2. The cytochrome P450 enzymes usually act as terminal oxidases in multicomponent electron transfer chains, called P450-containing monooxygenase systems. P450-containing monooxygenase systems primarily fall into two major classes: bacterial/mitochondrial (type I), and microsomal (type II). All P450 enzymes can be categorised into two main groups, the so-called B- and E-classes: P450 proteins of prokaryotic 3-component systems and fungal P450nor (CYP55) belong to the B-class; all other known P450 proteins from distinct systems are of the E-class. This family contains a number of subtypes of both B and E classes.

The disclosed NOV49 nucleic acid of the invention encoding a Leukotriene-B4 Omega- Hydroxylase-like protein includes the nucleic acid whose sequence is provided in Table 49A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 49A while still encoding a protein that maintains its Leukotriene-B4 Omega- Hydroxylase -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or

variant nucleic acids, and their complements, up to about 17 percent of the bases may be so changed.

The disclosed NOV49 protein of the invention includes the Leukotriene-B4 Omega-Hydroxylase-like protein whose sequence is provided in Table 49B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 49B while still encoding a protein that maintains its Leukotriene-B4 Omega-Hydroxylase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 25 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$ that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Leukotriene-B4 Omega-Hydroxylase -like protein (NOV49) is a member of a “Leukotriene-B4 Omega- Hydroxylase family”. Therefore, the NOV49 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV49 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Atherosclerosis, Aneurysm, Hypertension, Fibromuscular dysplasia, Stroke, Scleroderma, Obesity, Transplantation, Myocardial infarction, Embolism, Cardiovascular disorders, Bypass surgery, Osteoporosis, Hypercalcaemia, Arthritis, Ankylosing spondylitis, Scoliosis, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies, Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome, and/or other diseases and pathologies.

NOV49 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV49 substances for use in

therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV49 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV50

NOV50 includes three novel Protein Arginine N-Methyltransferase 2-like proteins disclosed below. The disclosed sequences have been named NOV50a and NOV50b.

NOV50a

A disclosed NOV50a nucleic acid of 1196 nucleotides (also referred to as CG56771-01) encoding a Protein Arginine N-Methyltransferase 2-like protein is shown in Table 50A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 13-15 and ending with a TGA codon at nucleotides 1068-1070. The start and stop codons are shown in bold in Table 50A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 50A. NOV50a nucleotide sequence (SEQ ID NO:187).

```

AGAGCGGCCAAGATGTCGCAGCCCAAGAAAAGAAAGCTTGAGTCGGGGGGCGGCGCCGAAGGAGGGGAGGGA
ACTGAAGAGGAAGATGGCGCGGAGCGGGAGGCGGCCCTGGAGCGACCCCGAGGACTAAGCGGGAACGGGAC
CAGCTGTACTACGAGTGCTACTCGGACGTTTCGGTCCACGAGGAGATGATCGCGGACCGCGTCCGCACCGAT
GCCTACCGCCTGGGTATCCTTCGGAACCTGGGCAGCACTGCGAGGCAAGACGGTACTGGACGTGGGCGCGGGC
ACCGGCATTCTGAGCATCTTCTGTGCCAGGCCGGGGCCCGCGCGTGTACGCGGTAGAGGCCAGCGCCATC
TGGCAACAGGCCCGGGAGGTGGTGGCGTTCAACGGGCTGGAGGACCGGGTGACGTCTGCCGGGACCAAGTG
GAGACTGTAGAGTTGCCGGAACAGGTGGATGCCATCGTGAGCGAGTGGATGGGCTACGGAACCTGTCACGAG
TCCATGCTGAGCTCCGTCCTCCACGCGCGAACCAGTGGCTGAAGGAGGGCGGTCTTCTCTGCCGGCCTCC
GCCGAGCTCTTCATAGCCCCCATCAGCGACAGATGCTGGAATGGCGCCTGGGCTTCTGGAGCCAGGTGAAG
CAGCACTATGGTGTGGACATGAGCTGCCTGGAGGGCTTCGCCACGCGTGTCTCATGGGCTAGAGCTCTCCC
GCGCCGCTTGGAGCAGGAGCTGGAGGCCGAGTGGGCGGGCGCTTCCGCTGCAGCTGCTATGGCTCGGCGC
CCATGCATGGCTTTGCCATCTGGTTCAGGTGACCTTCCCTGGAGGGGAGTCCGAGAAACCCCTGGTGTCTGT
CCACCTCGCCTTTTCACCCGCCCACTCACTGGAACAGGCGCTCCTCTACCTGAACGAGCCGGTGCAAGTGG
AGCAAGACACGGACGTTTCAGGAGAGATCACGCTGCTGCCCTCCCGGACAACCCCGTCGCCGTGCGCGTGC
TGCTGCGCTACAAAGTGGGAGACCAGGAGGAGAAGACAAAGACTTTGCCATGGAGGACTGAGCGTTGCCTT
TTCTCCAGCTACCTCCAAAGCAGCCTGACCTGCGTGGGAGAGGCCCACTCGGAGATCGTTGTGCAGGGA
TTGTCCGGCGAGGACGTGCTGGCCCGGCCGACGCGCTTTGCTCA

```

In a search of public sequence databases, the NOV50a nucleic acid sequence, located on chromosome 19, has 681 of 719 bases (94%) identical to a gb:GENBANK-
ID:AK001421|acc:AK001421.1 mRNA from *Homo sapiens* (cDNA FLJ10559 fis, clone NT2RP2002618, weakly similar to Protein Arginine N-Methyltransferase 2 (EC 2.1.1.-)) (E = 2.7e⁻¹³⁶).

The disclosed NOV50a polypeptide (SEQ ID NO:188) encoded by SEQ ID NO:187 has 375 amino acid residues and is presented in Table 50B using the one-letter amino acid

code. Signal P, Psort and/or Hydropathy results predict that NOV50a has no signal peptide and is likely to be localized to the nucleus with a certainty of 0.7000. Alternatively, NOV50a may also localize to the microbody (peroxisome) with a certainty of 0.2641, to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a
5 certainty of 0.1000.

Table 50B. Encoded NOV50a protein sequence (SEQ ID NO:188).

MSQPKKRKLESGGGAEGGECTEEEDGAEREAALERPRRTKRERDQLYYECYSDVSVHEEMIADRVRTDAYRL GILRNWAALRGKTVLDVGAGTGILSIFCAQAGARRVYAVEASAIWQQAREVVRFNGLEDVRHVLPGPVETVE LPEQVDAIVSEWMGYGLLHESMLSSVLHARTKWLKEGGLLPASAELFIAPISDQMLEWRLGFWSQVKQHYG VDMSCLEGFATRCLMGHSEIVVQGLSGEDVLARPQRFQAQLELSRAGLEQELEAGVGGRFCSCYGSAPMHGF AIWFQVTFPGGESEKPLVLSTSPFHPATHWKQALLYLNPEVQVEQDQTDVSGEITLLPSRDNPRLRLVLLRYK VGDQEEKTKDFAMED

A search of sequence databases reveals that the NOV50a amino acid sequence has 316
of 316 amino acid residues (100%) identical to, and 316 of 316 amino acid residues (100%)
10 similar to, the 316 amino acid residue ptnr:SPTREMBL-ACC:Q9NVR8 protein from *Homo sapiens* (Human) (CDNA FLJ10559 FIS, Clone NT2RP2002618, Weakly Similar To Protein Arginine N-Methyltransferase 2 (EC 2.1.1.-)) ($E = 1.7e^{-169}$).

NOV50a is predicted to be expressed in at least lung, bronchus, kidney. This
information was derived by determining the tissue sources of the sequences that were included
15 in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV50b

In the present invention, the target sequence identified previously, NOV50a, was
subjected to the exon linking process to confirm the sequence. PCR primers were designed by
20 starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in
25 silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra,
30 brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney,

lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV50b. This differs from the previously identified sequence (NOV50a) at aminoacid position 15 A->G.

A disclosed NOV50b nucleic acid of 1165 nucleotides (also referred to as CG56771-02) encoding a Protein Arginine N-Methyltransferase 2-like protein is shown in Table 50C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 4-6 and ending with a TGA codon at nucleotides 1129-1131. The start and stop codons are shown in bold in Table 50C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 50C. NOV50b nucleotide sequence (SEQ ID NO:189).	
<p><u>AAGATGTCGCAGCCCAAGAAAAGAAAGCTTGAGTCGGGGGGCGGCGGCGAAGGAGGGGAGGGAAGTGAAGAG</u> GAAGATGGCGCGGAGCGGGAGGCGGCCCTGGAGCGACCCCGGAGGACTAAGCGGGAACGGGACCAGCTGTAC TACGAGTGCTACTCGGACGTTTCGGTCCACGAGGAGATGATCGCGGACCCGCTCCGCACCGATGCCATCCGC CTGGGTATCCTTCGGAAGTGGGCGAGCTGCGAGGCAAGACGGTACTGGACGTGGGCGCGGGCACCGGCATT CTGAGCATCTTCTGTGCCAGGCCGGGGCCGGCGCGTGTACGCGGTAGAGGCCAGCGCCATCTGGCAACAG GCCCGGGAGGTGGTGCGGTCAACGGGCTGGAGGACCGGGTGCACGTCTGCCGGGACAGTGGAGACTGTA GAGTTGCCGAACAGGTGGATGCCATCGTGAGCGAGTGGATGGGCTACGGACTCCTGCACGAGTCCATGCTG AGCTCCGTCCTCCACGCGGAACCAAGTGGCTGAAGGAGGGCGGTCTTCTCCTGCCGGCCTCCGCCGAGCTC TTCATAGCCCCCATCAGCGACCAGATGCTGGAATGGCGCCTGGGCTTCTGGAGCCAGGTGAAGCAGCACTAT GGTGTGGACATGAGCTGCCTGGAGGGCTTCGCCACGCGCTGTCTCATGGGCCACTCGGAGATCGTTGTGTCAG GGATTGTCCGGCGAGGACGTGTGCCCGCGCGCAGCGCTTTGCTCAGCTAGAGCTCTCCCGCGCGGCTTG GAGCAGGAGCTGGAGGCCGAGTGGGCGGGCGCTTCCGCTGCAGCTGCTATGGCTCGGCGCCCATGCATGGC TTTGCCATCTGGTTCCAGGTGACCTTCCTTGGAGGGGAGTCCGAGAAACCCCTGGTGTGTCCACCTCGCCT TTTACCCGGCCACTCACTGGAACAGGCGCTCCTTACCTGAACGAGCCGGTGAAGTGGAGCAAGACACG GACGTTTCAGGAGAGATCACGCTGCTGCCCTCCCGGACAACCCCGTCCCTGCGCGTGTGTGCGCTAC AAAGTGGGAGACCAGGAGGAGAAGACCAAGACTTTGCCATGGAGGACTGAGCGTTGCCTTTTCCCCAGCT ACCTCCCAAAGCA</p>	

In a search of public sequence databases, the NOV50b nucleic acid sequence, located on chromosome 19, has 1090 of 1091 bases (99%) identical to a gb:GENBANK-ID:AK001421|acc:AK001421.1 mRNA from *Homo sapiens* (cDNA FLJ10559 fis, clone NT2RP2002618, weakly similar to Protein Arginine N-Methyltransferase 2 (EC 2.1.1.-)) (E = 2.2e⁻²⁴⁰).

The disclosed NOV50b polypeptide (SEQ ID NO:190) encoded by SEQ ID NO:189 has 375 amino acid residues and is presented in Table 50D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV50b has no signal peptide

and is likely to be localized to the nucleus with a certainty of 0.7000. Alternatively, NOV50b may also localize to the microbody (peroxisome) with a certainty of 0.2766, to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

5

Table 50D. Encoded NOV50b protein sequence (SEQ ID NO:190).
MSQPKRKLES GGGGEGEGTEEDGAEREALERPRTKRERDQLYYECYSDVSVHEEMIADRVRTDAYRL GILRNWAALRGKTVLDVGAGTGILSIFCAQAGARRVYAVEASAIWQQAREVVRFNGLIEDRVHVLPGPVETVE LPEQVDAIVSEWMGYGLLHESMLSSVLHARTKWLKEGGLLPASAELFIAPISDQMLEWRLGFWSQVKQHYG VDMSCLEGFATRCLMGHSEIVVQGLSGEDVLARPORFAQLELSRAGLEQELEAGVGGRFCRCYGSAPMHGF AIWFQVTFPGGESEKPLVLSTSPFHPATHWKQALLYLNEPVQVEQD TDVSGEITLLPSRDNPRLRLVLLRYK VGQDEEKTDFAMED

A search of sequence databases reveals that the NOV50b amino acid sequence has 316 of 316 amino acid residues (100%) identical to, and 316 of 316 amino acid residues (100%) similar to, the 316 amino acid residue ptnr:TREMBLNEW-ACC:AAH02729 protein from *Homo sapiens* (Human) (Hypothetical 35.2 Kda Protein) ($E = 1.8e^{-169}$).

10

NOV50b is predicted to be expressed in at least lung, bronchus, kidney. .

NOV50a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 50E

Table 50E. BLAST results for NOV50a					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15822652 gb AAK85733.1 (AY043278)	arginine methyltransferase 6 [<i>Homo sapiens</i>]	375	374/375 (99%)	374/375 (99%)	0.0
gi 8922515 ref NP_060607.1 (NM_018137)	hypothetical protein FLJ10559 [<i>Homo sapiens</i>]	316	316/316 (100%)	316/316 (100%)	0.0
gi 9293956 dbj BAB01859.1 (AP000383)	protein arginine N-methyltransferase-like protein[<i>Arabidopsis thaliana</i>]	399	148/317 (46%)	193/317 (60%)	5e-66
gi 15231011 ref NP_188637.1 (NC_003074)	arginine methyltransferase, putative [<i>Arabidopsis thaliana</i>]	409	143/310 (46%)	185/310 (59%)	1e-65
gi 15233606 ref NP_194680.1 (NC_003075)	arginine methyltransferase (pam1) [<i>Arabidopsis</i>]	390	135/365 (36%)	201/365 (54%)	4e-58

15

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 50F. In the ClustalW alignment of the NOV50 protein, as

well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 50F. ClustalW Analysis of NOV50

1)	Novel NOV50a	(SEQ ID NO:188)	
2)	Novel NOV50b	(SEQ ID NO:190)	
3)	gi 15822652 gb AAK85733.1	(AY043278) arginine methyltransferase 6 [<i>Homo sapiens</i>]	(SEQ ID NO:552)
4)	gi 8922515 ref NP_060607.1	(NM_018137) hypothetical protein FLJ10559 [<i>Homo sapiens</i>]	(SEQ ID NO:553)
5)	gi 9293956 dbj BAB01859.1	(AP000383) protein arginine N-methyltransferase-like protein[<i>Arabidopsis thaliana</i>]	(SEQ ID NO:554)
6)	gi 15231011 ref NP_188637.1	(NC_003074) arginine methyltransferase, putative [<i>Arabidopsis thaliana</i>]	(SEQ ID NO:555)
7)	gi 15233606 ref NP_194680.1	(NC_003075) arginine methyltransferase (pam1) [<i>Arabidopsis</i>]	(SEQ ID NO:556)

		10	20	30	40	50	60	
NOV50a	1	-----	MSQPKK	KLES	GGGAB	GGEG	---TEED	GA 27
NOV50b	1	-----	MSQPKK	KLES	GGGEG	GGEG	---TEED	GA 27
gi 15822652	1	-----	MSQPKK	KLES	GGGEG	GGEG	---TEED	GA 27
gi 8922515	1	-----	-----	-----	-----	-----	-----	1
gi 9293956	1	MQSGG	DFSNG	FHGDH	HRELE	EDKQ	GPSL	SSFGRAKKRSHAGARDPRGGLANVLRVSDQL 60
gi 15231011	1	MQSGG	DFSNG	FHGDH	HRELE	EDKQ	GPSL	SSFGRAKKRSHAGARDPRGGLANVLRVSDQL 60
gi 15233606	1	-----	MTKNS	NHDENE	FISFEP	NQNT	KIR-FE	DADEDEVAEGSGVAGEET--PQDESMF 51

		70	80	90	100	110	120	
NOV50a	28	EREAA	LERPR	RTK---	RERDQ	LYECY	SDVSV	HEEMIADRVRTDAYRLGILRNWAAALRGK 84
NOV50b	28	EREAA	LERPR	RTK---	RERDQ	LYECY	SDVSV	HEEMIADRVRTDAYRLGILRNWAAALRGK 84
gi 15822652	28	EREAA	LERPR	RTK---	RERDQ	LYECY	SDVSV	HEEMIADRVRTDAYRLGILRNWAAALRGK 84
gi 8922515	1	-----	-----	-----	-----	-----	-----	MIADRVRTDAYRLGILRNWAAALRGK 25
gi 9293956	61	GEHKS	LETSE	SSPPPC	TDFD	VAYE	HSYAH	VGHEEMIADRVRTDAYRLGILRNWAAALRGK 120
gi 15231011	61	GEHKS	LETSE	SSPPPC	TDFD	VAYE	HSYAH	VGHEEMIADRVRTDAYRLGILRNWAAALRGK 120
gi 15233606	52	DAGES	ADTA	EAVT	IDD--	TTSAD	YVFD	SYSHFGIHEEMIKDVRRTKTYQNVLYQNKFLIKDK 109

		130	140	150	160	170	180	
NOV50a	85	TVLDV	GAGTG	GILSIF	CAQAG	ARRVY	AVEASA	IWQOAREVVRFNGLGDRVHVLPCPVETVE 144
NOV50b	85	TVLDV	GAGTG	GILSIF	CAQAG	ARRVY	AVEASA	IWQOAREVVRFNGLGDRVHVLPCPVETVE 144
gi 15822652	85	TVLDV	GAGTG	GILSIF	CAQAG	ARRVY	AVEASA	IWQOAREVVRFNGLGDRVHVLPCPVETVE 144
gi 8922515	26	TVLDV	GAGTG	GILSIF	CAQAG	ARRVY	AVEASA	IWQOAREVVRFNGLGDRVHVLPCPVETVE 85
gi 9293956	121	VVVDV	GCGTG	GILSIF	CAQAG	AKRVY	AVDASD	IAVQAKEVVKANGLSDKVIVLHGRVEDVE 180
gi 15231011	121	VVVDV	GCGTG	GILSIF	CAQAG	AKRVY	AVDASD	IAVQAKEVVKANGLSDKVIVLHGRVEDVE 180
gi 15233606	110	IVLDV	GAGTG	GILSIF	CAQAG	AHVVY	AVECS	OMADMAKETVVKANGFSVITVTKGKIEEIE 169

		190	200	210	220	230	240	
NOV50a	145	LP-EQ	VDATV	SEWMG	YGLLH	ESML	SSVLH	ARTKWLKEGGILLPASAELEFIAPISD-OMLE 202
NOV50b	145	LP-EQ	VDATV	SEWMG	YGLLH	ESML	SSVLH	ARTKWLKEGGILLPASAELEFIAPISD-OMLE 202
gi 15822652	145	LP-EQ	VDATV	SEWMG	YGLLH	ESML	SSVLH	ARTKWLKEGGILLPASAELEFIAPISD-OMLE 202
gi 8922515	86	LP-EQ	VDATV	SEWMG	YGLLH	ESML	SSVLH	ARTKWLKEGGILLPASAELEFIAPISD-OMLE 143
gi 9293956	181	TD-EE	VDVIT	SEWMG	YMLLY	ESML	GSVIT	ARDRWLKPFGGLILPSSHATLYMAPISHPDRYS 239
gi 15231011	181	TD-EE	VDVIT	SEWMG	YMLLY	ESML	GSVIT	ARDRWLKPFGGLILPSSHATLYMAPISHPDRYS 239
gi 15233606	170	LP-TP	KVDVIT	SEWMG	YELL	FENML	DSVL	YARDKWLVEGGVLLPDKASHHTAILESEYKE 229

			250	260	270	280	290	300	
	NOV50a	203	WRLGFWSQVKQHYGVDMSCLEGFATRCIMGHSEIVVQGLSGEDVLARPQRFQAELELS	259					
	NOV50b	203	WRLGFWSQVKQHYGVDMSCLEGFATRCIMGHSEIVVQGLSGEDVLARPQRFQAELELS	259					
5	gi 15822652	203	WRLGFWSQVKQHYGVDMSCLEGFATRCIMGHSEIVVQGLSGEDVLARPQRFQAELELS	259					
	gi 8922515	144	WRLGFWSQVKQHYGVDMSCLEGFATRCIMGHSEIVVQGLSGEDVLARPQRFQAELELS	200					
	gi 9293956	240	HSTDFWRNV---YGITDMSAMMQLAKQCAF--EEPSVESISGENVLTWPEVV	285					
	gi 15231011	240	HSTDFWRNV---YGITDMSAMMQLAKQCAF--EEPSVESISGENVLTWPEVVVKHIDCKTIK	294					
10	gi 15233606	230	DKIEFWNSV---YGITDMSCLIKKA---M--MEPEVDITVDQNQIVTDSRLKTKMDIS	277					
			310	320	330	340	350	360	
	NOV50a	259	-----RAG-----LEQLEAGVGGRFRCSCYGSAPMHGFATWFOVT	295					
	NOV50b	259	-----RAG-----LEQLEAGVGGRFRCSCYGSAPMHGFATWFOVT	295					
15	gi 15822652	259	-----RAG-----LEQLEAGVGGRFRCSCYGSAPMHGFATWFOVT	295					
	gi 8922515	200	-----RAG-----LEQLEAGVGGRFRCSCYGSAPMHGFATWFOVT	236					
	gi 9293956	285	-----FPLEFSGPASSPAKNTSETSIASGSSSISPSGE	318					
	gi 15231011	295	IQELDSVTARYKFNSMMRAPMHGFATWFOVDFSGPASSPAKNTSETSIASGSSSISPSGE	354					
20	gi 15233606	277	-----KMSSGDAFTAPFKLVAQRNDYTHALVAYEDVS	310					
			370	380	390	400	410	420	
	NOV50a	296	FPGGESEKP---LVLSTSPFHHPATHWKQALLYLNEPVQVEQDITDVSGETITLLPSRDNP	352					
	NOV50b	296	FPGGESEKP---LVLSTSPFHHPATHWKQALLYLNEPVQVEQDITDVSGETITLLPSRDNP	352					
25	gi 15822652	296	FPGGESEKP---LVLSTSPFHHPATHWKQALLYLNEPVQVEQDITDVSGETITLLPSRDNP	352					
	gi 8922515	237	FPGGESEKP---LVLSTSPFHHPATHWKQALLYLNEPVQVEQDITDVSGETITLLPSRDNP	293					
	gi 9293956	319	VNKKRTNPSDALVLTSPSPFTHWQOTLVYFYDPIQVEQDQVLEGSVTLSSQSKENKRF	378					
	gi 15231011	355	VNKKRTNPSDALVLTSPSPFTHWQOTLVYFYDPIQVEQDQVLEGSVTLSSQSKENKRF	392					
30	gi 15233606	311	FTMCHKLLG----FSTGPKSRATHWKQTVLYLEDVLTICEGETITCTMSVSENKKNPRD	365					
			430	440					
	NOV50a	353	LRVLLRYKVGDOEEKTKDFAMED--	375					
	NOV50b	353	LRVLLRYKVGDOEEKTKDFAMED--	375					
35	gi 15822652	353	LRVLLRYKVGDOEEKTKDFAMED--	375					
	gi 8922515	294	LRVLLRYKVGDOEEKTKDFAMED--	316					
	gi 9293956	379	MNIHLEYSSAGRSFVKESVMR----	399					
	gi 15231011	392	---TNEIISNSASAYIIEKN-----	409					
40	gi 15233606	366	TDKLSYSYLNQCHCKISRTQHYKMR	390					

Table 49G lists the domain description from DOMAIN analysis results against NOV50. This indicates that the NOV50 sequence has properties similar to those of other proteins known to contain this domain.

Table 50G Domain Analysis of NOV50

gnl|Pfam|pfam01209, Ubie_methyltran, ubiE/COQ5 methyltransferase family. (SEQ ID NO:849)
CD-Length = 237 residues, 63.3% aligned
Score = 35.0 bits (79), Expect = 0.008

NOV50:	40	KRERDQLYYECYSYSDVSVHEEMIADRVRTDAYRLGILRNW-----AALRGKTVLDVG	90
Sbjct:	5	KEEKEQKVHHVFASVAKKYDLM----NDVMSFGIHLRWKDHFTMKLMGPKRGKKFLDVA	59
NOV50:	91	AGTGILSI-FCAQAGAR-RVYAVEASA-IWQQAREVVRFNGLEDVRHVLPGPVETVELPE	147
Sbjct:	60	GGTGDIARLLRHAGTSGKVVVLDINENMLKVGKKRAE-EEGKIRIEWLCANAE--ELPF	116
NOV50:	148	QVDAIVSEWMGYGLLHESMLSSSVLHARTKWLKEGGLLL	185

Methyl transfer from S-Adenosyl-L-methionine (SAM) to either nitrogen, oxygen or carbon atoms is frequently employed in diverse organisms ranging from bacteria to plants and mammals. The reaction is catalyzed by methyltransferases (MTases) and modifies DNA, RNA, proteins and small molecules like catechol. The catalytic domain of SAM-MTases is of the alpha/beta type with a central mixed beta-sheet around which several alpha-helices are arranged. Topologically it can be divided into two halves. The first half, formed by beta1-alphaA-beta2-alphaB-beta3-alphaC, is mainly responsible for SAM binding. The second half, beta4-alphaD-beta5-alphaE-beta6-beta7, is primarily responsible for catalysis. According to the sequential order of these two sites, the SAM-MTases can be divided into three families. Protein arginine methylation has been implicated in signal transduction, nuclear transport and transcription regulation. Protein arginine methyltransferases (PRMTs) mediate the AdoMet-dependent methylation of many proteins, including many RNA binding proteins involved in various aspects of RNA processing and/or transport.

The bulk of methylated arginine residues in eukaryotic cells are found in heterogeneous nuclear ribonucleoproteins (hnRNPs), RNA-binding proteins that play essential roles in the metabolism of nuclear pre-mRNA. Lin et al. (1996) identified a rat cDNA encoding PRMT1 (protein-arginine N-methyltransferase 1; EC 2.1.1.23). Recombinant PRMT1 methylated histones and hnRNPA1 (164017) in vitro. By using a yeast 2-hybrid screen to identify proteins that interact with the intracytoplasmic domain of the interferon-alpha/beta receptor-1 (IFNAR1; 107450), Abramovich et al. (1997) identified a human cDNA encoding a protein that was nearly identical to PRMT1. The deduced 361-amino acid protein was designated IR1B4 for 'interferon receptor-1-bound protein 4.' Epitope-tagged IR1B4 bound the IFNAR1 intracytoplasmic domain in vitro. Antibodies against IFNAR1 coimmunoprecipitated a methyltransferase activity from human cell extracts. An antisense oligonucleotide strongly reduced methyltransferase activity in human cells, and caused them to become more resistant to growth inhibition by interferon. Abramovich et al. (1997) concluded that protein methylation, like phosphorylation, may be an important signaling mechanism for certain cytokine receptors. Scott et al. (1998) identified HRMT1L2 transcripts with variable 5-prime ends that encode 3 protein variants with different N-terminal regions. Rat PRMT1 and HRMT1L2 variant 2 (v.2) share 95% sequence identity, but diverge at their N termini. The amino acid sequences of HRMT1L2 and HRMT1L1 (601961) are 27% identical. Recombinant protein methylated human hnRNPA1 and a yeast hnRNP in vitro. The HRMT1L2 gene complemented mutations in the yeast hnRNP methyltransferase gene HMT1.

Northern blot analysis revealed that HRMT1L2 is expressed as a predominant 1.4-kb mRNA in various adult and fetal tissues. Additional larger and smaller bands were observed in some tissues.

5 The disclosed NOV50 nucleic acid of the invention encoding a Protein Arginine N-Methyltransferase 2-like protein includes the nucleic acid whose sequence is provided in Table 50A, 50C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 50A or 50C while still encoding a protein that maintains its Protein Arginine N-Methyltransferase 2-like activities and physiological functions, or a fragment of such a nucleic acid. The invention
10 further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar
15 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 6 percent of the bases may be so changed.

20 The disclosed NOV50 protein of the invention includes the Protein Arginine N-Methyltransferase 2-like protein whose sequence is provided in Table 50B or 50D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 50B or 50D while still encoding a protein that maintains its Protein Arginine N-Methyltransferase 2-like activities and physiological
25 functions, or a functional fragment thereof. In the mutant or variant protein, up to about 63 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

30 The above disclosed information suggests that this Protein Arginine N-Methyltransferase 2-like protein (NOV50) is a member of a "Protein Arginine N-Methyltransferase 2 family". Therefore, the NOV50 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target,

In a search of public sequence databases, the NOV51 nucleic acid sequence, located on chromosome 22, has 967 of 984 bases (98%) identical to a gb:GENBANK-ID:AP000534|acc:AP000534.1 mRNA from *Homo sapiens* (genomic DNA, chromosome 22q11.2, Cat Eye Syndrome region, clone:c23H5) ($E = 2.9e^{-208}$).

5 The disclosed NOV51 polypeptide (SEQ ID NO:192) encoded by SEQ ID NO:191 has 315 amino acid residues and is presented in Table 51B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV51 has a signal peptide and is likely to be localized to the plasma mambrane with a certainty of 0.6000. Alternatively, NOV51 may also localize to the Golgi body with a certainty of 0.4000, to the endoplasmic
10 reticulum (membrane) with a certainty of 0.3000, or to the microbody (peroxisome) with a certainty of 0.3000. The most likely cleavage site for NOV51 is between positions 40 and 41: IYA-LT.

Table 51B. Encoded NOV51 protein sequence (SEQ ID NO:192).

MNVSEPNSSFAFVNEFILQGFSCWEVTIQIFLFLSLETTIYALTITGNGAIAFALWCDRLHTPMYMFGLGDFS LEIWYVFSTVPKMLVNFLSEKTNISFAGCFLQYFFFFSLGTSECLLLTVMAFDQYLAI CRPLHYPNIMTGHL CAKLVLWCWCGFLWFLIPIVLISQMPFCGPNI IDHVVCDPGLFALDCVSAPRIQLFCYTLNSLVI FGNFL FI IGSYTIVLKAVLGTPSSTGRHKAFSTCGSHLAVVSPCYGLMVMYVSPGLGHSTGMQKI VTLFYAMVTPL FNPLIYSLQNKEIKAALRKVLGSSNI I

15 A search of sequence databases reveals that the NOV51 amino acid sequence has 191 of 314 amino acid residues (60%) identical to, and 226 of 314 amino acid residues (71%) similar to, the 324 amino acid residue ptmr:SPTREMBL-ACC:Q9WU86 protein from *Mus musculus* (Mouse) (Odorant Receptor S1) ($E = 7.0e^{-100}$).

NOV51 is predicted to be expressed in at least Apical microvilli of the retinal pigment
20 epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma
25 cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that
30 were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV51 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 51C

Table 51C. BLAST results for NOV51					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 15293807 gb AAK95096.1 (AF399611)	olfactory receptor [<i>Homo sapiens</i>]	217	205/217 (94%)	208/217 (95%)	4e-99
gi 9938010 ref NP_064684.1 (NM_020288)	odorant receptor S1 gene [<i>Mus musculus</i>]	324	191/314 (60%)	226/314 (71%)	5e-90
gi 17476501 ref XP_063251.1 (XM_063251)	similar to OLFACTORY RECEPTOR-LIKE PROTEIN F6 (H. sapiens) [<i>Homo sapiens</i>]	1056	134/293 (45%)	181/293 (61%)	2e-66
gi 15293805 gb AAK95095.1 (AF399610)	olfactory receptor [<i>Homo sapiens</i>]	217	142/217 (65%)	163/217 (74%)	2e-65
gi 17476700 ref XP_063315.1 (XM_063315)	similar to odorant receptor S1 gene (H. sapiens) [<i>Homo sapiens</i>]	195	131/189 (69%)	156/189 (82%)	2e-64

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 51D. In the ClustalW alignment of the NOV51 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 51D. ClustalW Analysis of NOV51

- 1) Novel NOV51 (SEQ ID NO:192)
- 2) gi|15293807|gb|AAK95096.1| (AF399611) olfactory receptor [*Homo sapiens*] (SEQ ID NO:557)
- 3) gi|9938010|ref|NP_064684.1| (NM_020288) odorant receptor S1 gene [*Mus musculus*] (SEQ ID NO:558)
- 4) gi|17476501|ref|XP_063251.1| (XM_063251) similar to OLFACTORY RECEPTOR-LIKE PROTEIN F6 (H. sapiens) [*Homo sapiens*] (SEQ ID NO:559)
- 5) gi|15293805|gb|AAK95095.1| (AF399610) olfactory receptor [*Homo sapiens*] (SEQ ID NO:560)
- 6) gi|17476700|ref|XP_063315.1| (XM_063315) similar to odorant receptor S1 gene (H. sapiens) [*Homo sapiens*] (SEQ ID NO:561)

NOV51 1 10 20 30 40 50 60 1

5	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	1	MPVLLPVHFSAKCPLLLLCDPANPPSEPLPSQGCFFIHRVLLDLSTAGESGNTAGFICD	60
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
10	NOV51	1	70 80 90 100 110 120	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	61	QALLTSPVREDGAENGLGFHQPVELHICGDAVGFGVGMGQRRKPMSPVWSPKISEKCASD	120
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
15	NOV51	1	130 140 150 160 170 180	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	121	TWCTDATYHREHSKPSGPGWEHGPLKPFEDWVPALPYPLWPQELLHCGSQSGDCMCLLLE	180
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
20	NOV51	1	190 200 210 220 230 240	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	181	SSRRSPPTLPPIPLTFPRLCQSFPLLTASGKEPSCGFTSALRRLYGCGAAERPQSPVTPKT	240
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
25	NOV51	1	250 260 270 280 290 300	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	241	ETSEQGPKDPPIHLAHPSDRALSPSCFLSLRAVILTCKNRDAQVEEGHREPPVLDCGYQ	300
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
30	NOV51	1	310 320 330 340 350 360	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	301	RSGTRGNHTRRICSTLRGSRIEAWAAATLQRGPHYFRKQQLGKDSWSVAEDWIEAFMLA	360
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
35	NOV51	1	370 380 390 400 410 420	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	361	FGVRVLWDASMALEAQRDPSSNDTKGKDQLTKRDQRNPQNPFALLQSAASDWSNPVCCR	420
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
40	NOV51	1	430 440 450 460 470 480	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
45	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
50	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
55	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
60	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
65	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1
70	NOV51	1	490 500 510 520 530 540	1
	gi 15293807	1	-----	1
	gi 9938010	1	-----	1
	gi 17476501	421	GYLTCASASLGEISSPHFPVHLNAPKCHWGLSSSPVERWMLRERKAVTDESSSSWMVAIR	480
	gi 15293805	1	-----	1
	gi 17476700	1	-----	1

446

			970	980	990	1000	1010	1020	
	NOV51	315	315
5	gi 15293807	217	-----	-----	-----	-----	-----	-----	217
	gi 9938010	324	-----	-----	-----	-----	-----	-----	324
	gi 17476501	961	HPAAGSPRDSRKVNVRVQKDPRRSVPKVETFISGSGPSCVGQCTGRVCILKGTRTISGGL						1020
	gi 15293805	217	-----	-----	-----	-----	-----	-----	217
10	gi 17476700	195	-----	-----	-----	-----	-----	-----	195
			1030	1040	1050				
	NOV51	315			315
15	gi 15293807	217	-----	-----	-----	-----			217
	gi 9938010	324	-----	-----	-----	-----			324
	gi 17476501	1021	WLEDPRKTRTTDFTHRKIKVTAGLAGEKVEPTLPRC						1056
	gi 15293805	217	-----	-----	-----	-----			217
	gi 17476700	195	-----	-----	-----	-----			195

Table 51E lists the domain descriptions from DOMAIN analysis results against NOV51. This indicates that the NOV51 sequence has properties similar to those of other proteins known to contain this domain.

Table 51E Domain Analysis of NOV51

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:810)
CD-Length = 254 residues, 100.0% aligned
Score = 105 bits (262), Expect = 4e-24

NOV51:	45	GNGAIAFALWCDRRLLHTPMYMFGLGDFSLEIYWVFSTVPKMLVNFLSEKTNISFAGCFLO	104
Sbjct:	1	GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLV	60
NOV51:	105	FYFFFSLGTSECLLLTVMAFDQYLAICRPLHYPNIMTGHLCAKLVILCWVCGFLWFLIPI	164
Sbjct:	61	GALFVVNGYASILLTAISIDRYLAIVHPLRYRRIPTPRRAKVLILLVWVLALLSLPPL	120
NOV51:	165	VLISQMPFCGPNIIDHVVCDPGLFALDCVSAPRIQLFCYTLNSLVIFGNFLFIIGSYTI	224
Sbjct:	121	LFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVPLLVILVCYTRILRTLKRA	180
NOV51:	225	VLKAVLGTPSSTGRHKAFSTCGSHLAVVSPCYGSLMVMYVSP----GLGHSTGMQKIVTL	280
Sbjct:	181	RSQRSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLDLSLCLLSIWRVLPALLITL	240
NOV51:	281	FYAMVTPLFNPLIY	294
Sbjct:	241	WLAYVNSCLNPIIY	254

G-Protein Coupled Receptor (GPCRs) have been identified as an extremely large family of protein receptors in a number of species. At the phylogenetic level they can be classified into four major subfamilies. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors. They are likely to be involved in the recognition and transduction of various signals mediated by G-Proteins, hence their name

G-Protein Coupled Receptors. The human GPCR genes are generally intron-less and belong to four gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

Olfactory receptors (ORs) have been identified as extremely large family of GPCRs in a number of species. As members of the GPCR family, these receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals. Like GPCRs, the ORs they can be expressed in a variety of tissues where they are thought to be involved in recognition and transmission of a variety of signals. The human OR genes are typically intron-less and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium.

The disclosed NOV51 nucleic acid of the invention encoding a Olfactory Receptor-like protein includes the nucleic acid whose sequence is provided in Table 51A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 51A while still encoding a protein that maintains its Olfactory Receptor -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2 percent of the bases may be so changed.

The disclosed NOV51 protein of the invention includes the Olfactory Receptor-like protein whose sequence is provided in Table 51B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 51B while still encoding a protein that maintains its Olfactory Receptor-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 54 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Olfactory Receptor -like protein (NOV51) is a member of a "Olfactory Receptor family". Therefore, the NOV51 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential

5 therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

10 The NOV51 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in developmental diseases, MHCII and III diseases (immune diseases), Taste and scent detectability Disorders, Burkitt's lymphoma, Corticoneurogenic disease, Signal Transduction pathway disorders, Retinal diseases including those involving photoreception, Cell Growth rate disorders; Cell Shape disorders, Feeding

15 disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,

20 hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian atrophy (DRPLA) Hypophosphatemic rickets,

25 autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders of the like.. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the OR -like protein may be useful in gene therapy, and the

30 OR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia,

asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary Osteodystrophy, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome, and/or other diseases and pathologies.

NOV51 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV51 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV51 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV52

A disclosed NOV52 nucleic acid of 3828 nucleotides (also referred to as CG56731-01) encoding a H326-like protein is shown in Table 52A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 177-179 and ending with a TAA codon at nucleotides 1968-1970. The start and stop codons are shown in bold in Table 52A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 52A. NOV52 nucleotide sequence (SEQ ID NO:193).

CTCTTAGCGCTCAGGTCTTTCTTCCGCCGACCCGAAGTCATCGCTGGGAGTACTGGTTGCCCTTTCTCA
GTCCTTCAGTGAATCTACAGAGCCTATTTCTCAGGAGCCTCAGCCTGGTCCTTACTTCAGTGATAAAAGGA
GGAAAGGCTGGCTACAGCAAACATCATTCAAGATGTCCAGCAAAGGGAGCAGCACAGATGGCAGAACAGACT
TAGCTAATGGAAGCCTGTCTAGCAGTCCAGAGGAGATGTCTGGAGCTGAAGAGGGGAGGGAGACATCCTCAG
GCATTGAAGTGGAGGCCTCAGACCTGAGTTTGAGCTTGACTGGGGATGATGGTGGCCCCAACCGCACCCAGCA
CAGAAAGTCGAGGCACAGACACAGAGAGCTCAGGTGAAGATAAGGACTCTGACAGCATGGAGGACACTGGTC
ATTACTCCATTAATGATGAAAATCGAGTCCATGACCGCTCAGAGGAAGAGGAAGAGGAGGAAGAAGAGGAGG
AAGAAGAGCAGCCTCGGCGCGGTGTACAGCGCAAGCGGGCTAACCGTGACCAGGACTCATCAGATGATGAGC
GGGCCCTAGAGGACTGGGTGTCTCAGAAACATCAGCTCTACCCCGACCTCGCTGGCAAGCCCTCCCTGCC
TTCCGGAGCGGGAGCTGGGTTCAAGTGCCCGCTTTGTCTATGAGGCCTGTGGGGCAAGAGTCTTTGTGCAGC
GTTTCCGCCTGCAGCATGGGCTTGAGGGCCATACTGGTTGTGTCAATACCCTGCACCTTTAACCAGCGCGGCA
CCTGGCTGGCCAGTGGCAGCGATGACCTGAAGTGGTGGTGTGGGATGGGTACGGCGGCAGCCAGTACTGG
ACTTTGAGAGTGGCCACAAAAGTAATGTGTCTCAGGCCAAGTTTCTTCTCAACAGTGGTGATTTCTACTTGG
CCATGTGTGCCCGTGACGGGCAGGTTTCGAGTAGCAGAACTGTCTGCCACACAGTGTGCAAGAAATACAAAC
GTGTGGCCACAGCACAAGGGAGCGTCCCAAGTTGGCACTGGAACCAGACTCTCCCTGTACGTTCTTATCTG
CAGTGAAGATGCAGTTGTTTTCACCATTGACCTGAGACAAAGACCGCCAGCGTCGAAACTGGTGGTGACAA
AAGAGAAAGAGAAGAAAGTGGGGCTATACGATCTATGTGAATCCTGCCAATACCCACCAGTTTGCAGTGG
GTGGACGAGATCAGTTTGTAAAGATTATGACCAGAGGAAATTGATGAGAATGAGAACAATGGAGTACTCA
AGAAGTTCTGTCTCATCAGCTGGTGAACAGTGAGTCCAAAGCAAACATCACCTGTCTTGTGTACAGCCAGC
ACGGCACAGAGCTCCTGGCCAGTTACAATGATGAAGACATTACCTCTTCAACTCCTCTCACAGTGATGGGG
CCCAGTATGTTAAGAGATACAAGGGCCACAGAAATAATGCCACAGTAAAGGCGTCAATTTCTATGGCCCCA
AGAGTGAGTTTGTGGTGAGCGGTAGTGACTGTGGGCACATCTTCTCTGGGAGAAATCATCTGCCAGATTA

A search of sequence databases reveals that the NOV52 amino acid sequence has 588 of 597 amino acid residues (98%) identical to, and 589 of 597 amino acid residues (98%) similar to, the 597 amino acid residue ptnr:SPTREMBL-ACC:Q12839 protein from *Homo sapiens* (Human) (H326) (E = 0.0).

NOV52 is predicted to be expressed in at least adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Amnion, Appendix, Bone, Bronchus, Brown adipose, Cervix, Chorionic Villus, Colon, Coronary Artery, Dermis, Epidermis, Foreskin, Hair Follicles, Hypothalamus, Kidney Cortex, Liver, Lung, Lung Pleura, Lymph node, Lymphoid tissue, Muscle, Ovary, Oviduct/Uterine Tube/Fallopian tube, Parathyroid Gland, Parotid Salivary glands, Peripheral Blood, Respiratory Bronchiole, Retina, Right Cerebellum, Skin, Synovium/Synovial membrane, Temporal Lobe, Thymus, Tonsils, Umbilical Vein, Urinary Bladder, Vein, Vulva, Whole Organism.

This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in plasma cells (myeloma) because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:HSU06631|acc: U06631.1) a closely related Human (H326) mRNA, complete cds homolog.

NOV52 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 52C

Table 52C. BLAST results for NOV52					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 13636682 ref XP_010501.2 (XM_010501)	similar to H326 (H. sapiens) [Homo sapiens]	597	597/597 (100%)	597/597 (100%)	0.0
gi 7657148 ref NP_056541.1 (NM_015726)	H326 [Homo sapiens]	597	588/597 (98%)	589/597 (98%)	0.0
gi 17485807 ref XP_066683.1 (XM_066683)	similar to H326 (H. sapiens) [Homo sapiens]	779	401/603 (66%)	463/603 (76%)	0.0

gi 17485821 ref XP_066690.1 (XM_066690)	similar to H326 (H. sapiens) [Homo sapiens]	577	387/601 (64%)	449/601 (74%)	0.0
gi 6679281 ref NP_032847.1 (NM_008821)	plasmacytoma expressed transcript 2 [Mus musculus]	747	308/469 (65%)	383/469 (80%)	0.0

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 52D. In the ClustalW alignment of the NOV52 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 52D. ClustalW Analysis of NOV52

- 1) Novel NOV52 (SEQ ID NO:194)
- 2) gi|13636682|ref|XP_010501.2| (XM_010501) similar to H326 (H. sapiens) [Homo sapiens] (SEQ ID NO:562)
- 3) gi|7657148|ref|NP_056541.1| (NM_015726) H326 [Homo sapiens] (SEQ ID NO:563)
- 4) gi|17485807|ref|XP_066683.1| (XM_066683) similar to H326 (H. sapiens) [Homo sapiens] (SEQ ID NO:564)
- 5) gi|17485821|ref|XP_066690.1| (XM_066690) similar to H326 (H. sapiens) [Homo sapiens] (SEQ ID NO:565)
- 6) gi|6679281|ref|NP_032847.1| (NM_008821) plasmacytoma expressed transcript 2 [Mus musculus] (SEQ ID NO:566)

			10	20	30	40	50	60	
	NOV52	1						1
25	gi 13636682	1	-----						1
	gi 7657148	1	-----						1
	gi 17485807	1	MANSKNPVEIEIMVFSSTLNMPDVGDNKPLSIKQVNDISRQSTMNSVHLLNLLLCFATAL						60
	gi 17485821	1	-----						1
30	gi 6679281	1	--MSSHESYTNAAETPENISILSCLGETSGALVDTKTISDIKTMDPRVSLTP-----						50
			70	80	90	100	110	120	
	NOV52	1						1
35	gi 13636682	1	-----						1
	gi 7657148	1	-----						1
	gi 17485807	61	KHHGLGDLQAACTLRQRVDVSLIGE QHPTRI PVI IGYKSEEQLPVLDKTKFLVLANQAFF						120
	gi 17485821	1	-----						1
	gi 6679281	51	SSDVTGTEDSSVLTPQSTDVN-----SVDSYQGYEGDDDDDEDDDDKDGDSNLPSL						102
40			130	140	150	160	170	180	
	NOV52	1						1
	gi 13636682	1	-----						1
	gi 7657148	1	-----						1
45	gi 17485807	121	LLRCMRSEKDEDDGFLYTNIFQGAASFDLQKPAATHSVRCLSGWGS GPGKTRAVWMLRGGE						180
	gi 17485821	1	-----						1
	gi 6679281	103	EDSDNFISCLSENSIYPQNVENGEVVEEQSLGRRFHPYELAG-----EVVEGQ						150
50			190	200	210	220	230	240	
								

5

10

15

20

25

30

35

40

45

50

55

60

65

70

```

NOV52      1  -----MSSKGSSTDGRTDLANGSLSSSPPEMSGAAEGRE 34
gi|1363682| 1  -----MSSKGSSTDGRTDLANGSLSSSPPEMSGAAEGRE 34
gi|7657148| 1  -----MSSKGSSTDGRTDLANGSLSSSPPEMSGAAEGRE 34
gi|17485807| 181 QANFLPRAFRLGLAFVDPDLLQOTSFKMSHQEGSDTGLPDLGTESLSSSPPEMSGAAVATE 240
gi|17485821| 1  -----MSHQEGSTGGLPDLVTESLSSSPPEMSGVAAVTA 34
gi|6679281| 151 GGGSLFPYPYELEAGEVVEAQNVQNLFHRYLEEEGEVVEAQVVOFMFPYYELEAGEVVEAE 210

          250      260      270      280      290      300
NOV52      35  TSS---GIBVEASDLSLSLTGD-----DGG-PNRTSTESRGTDTESSGEDKDSDSM 81
gi|1363682| 35  TSS---GIBVEASDLSLSLTGD-----DGG-PNRTSTESRGTDTESSGEDKDSDSM 81
gi|7657148| 35  TSS---GIBVEASDLSLSLTGD-----DGG-PNRTSTESRGTDTESSGEDKDSDSM 81
gi|17485807| 241 ASS---DIDTIATSELSVTVTGDSGSD---SRDGGFPNDASTENRSSDOESASEDIELEST 293
gi|17485821| 35  ASS---DIDMAATEPS---TGDGGD---TRDGGFLNDASTENQNTDESSESEDVELESM 84
gi|6679281| 211 EVQGGFFQRYVELEAREVICAQGGQGLSRHYGLEGGGEVVEATAVRRLTQHHELEGEDVDVDDQ 270

          310      320      330      340      350      360
NOV52      82  EDTGHSYINDENRVHDRSEEEEEEEEEEEEEEQPRRRVQRKRANRDQSSDDDERALEDWVS 141
gi|1363682| 82  EDTGHSYINDENRVHDRSEEEEEEEEEEEEEEQPRRRVQRKRANRDQSSDDDERALEDWVS 141
gi|7657148| 82  EDTGHSYINDENRVHDRSEEEEEEEEEEEEEEQPRACTAQAAANRDQSSDDDERALEDWVS 141
gi|17485807| 294 ED-----FEHFLMSGSGGNHEQYSLEEDQALEEWVS 324
gi|17485821| 85  GEG-----LFGYPLPRMPCRCGGTNHDCCLLDEDALEEWVS 121
gi|6679281| 271 EG-----SSEMHEETSSEDSSQVDIEDDSLIDEWIA 301

          370      380      390      400      410      420
NOV52      142 SETSALPRPRWQALPALRRERELGSSARFVYEACGARVFVQFRRLQHCLEGHTGCVNTLHF 201
gi|1363682| 142 SETSALPRPRWQALPALRRERELGSSARFVYEACGARVFVQFRRLQHCLEGHTGCVNTLHF 201
gi|7657148| 142 SETSALPRPRWQALPALRRERELGSSARFVYEACGARVFVHGFRRLQHCLEGHTGCVNTLHF 201
gi|17485807| 325 SETSALPRPRWQVVTALRQRLGSRERFVYEACGARAFVQFRRLQKRLADHVGCVNTVHF 384
gi|17485821| 122 SETSALPRSRWQVLTALRQRLGSSARFVYEACGARTFVQFRRLQKLLGSHAGSVSTIHF 181
gi|6679281| 302 LETSPLPRPRWNVLSALRRDRQLGSSGRFVYEACGARLFVQFRFLSEHVEGHSVCVNTVHF 361

          430      440      450      460      470      480
NOV52      202 NQRGTWLASGSDDLKVVVVDWVRRQPVLDLFESGHKSNVFOAKFLPNSGDSTLAMCARDGQ 261
gi|1363682| 202 NQRGTWLASGSDDLKVVVVDWVRRQPVLDLFESGHKSNVFOAKFLPNSGDSTLAMCARDGQ 261
gi|7657148| 202 NQRGTWLASGSDDLKVVVVDWVRRQPVLDLFESGHKSNVFOAKFLPNSGDSTLAMCARDGQ 261
gi|17485807| 385 NQRGTRLASSGDDLKVIIVDWVRQRPVLNLFESGHTNNVFOAKFLPNSGDSTLAMCARDGQ 444
gi|17485821| 182 NQRGTRLASSGDDLRLVIIVDWVRQRPVLNLFESGHGDIVNFOAKFFPNCGDSTLAMCGHDGQ 241
gi|6679281| 362 NQHGTLLASGSDDLKVIIVDWLKKRSVLNFDLFESGHKNNILFOAKFLPNCNDAILAMCGRDGQ 421

          490      500      510      520      530      540
NOV52      262 VRVAELSATQCCKNTKRVAQHKGASHKLALEPDSPTFLSAGEDAVVFTIDLRQDRPASK 321
gi|1363682| 262 VRVAELSATQCCKNTKRVAQHKGASHKLALEPDSPTFLSAGEDAVVFTIDLRQDRPASK 321
gi|7657148| 262 VRVAELSATQCCKNTKRVAQHKGASHKLALEPDSPTFLSAGEDAVVFTIDLRQDRPASK 321
gi|17485807| 445 VRVAELINASYFNNTKVAQHGRPAHKLALPDSPIYKFLTSGEDAVVFTIDLRQDRPASK 504
gi|17485821| 242 VRVAELINASYCENTKRVAKHGRPAHKLALPDSPIYKFLTSGEDAVVFTIDLRQDRPASK 301
gi|6679281| 422 VRVAQLSAVAGTHMTKRLVHKHGASHRLGLEPDSPFRLFTSGEDAVVENIDLRQAHPAK 481

          550      560      570      580      590      600
NOV52      322 LVVTKEKEKKVGLYTIIVNPANTHQFVAGGQDQFVRIYDQRKIDENENNGVLKKFCPHHL 381
gi|1363682| 322 LVVTKEKEKKVGLYTIIVNPANTHQFVAGGQDQFVRIYDQRKIDENENNGVLKKFCPHHL 381
gi|7657148| 322 LVVTKEKEKKVGLYTIIVNPANTHQFVAGGQDQFVRIYDQRKIDENENNGVLKKFCPHHL 381
gi|17485807| 505 VVVTRENDKKVGLYTIIVNPANTYQFVAGGQDQFVRIYDQRKIDKKENNGVLKKFTPHHL 564
gi|17485821| 302 VVVTRENDKKVGLYTIIVNPANIYQFVAGGQDQFVRIYDQRRIIDKKENNGVLKKFTPHHL 361
gi|6679281| 482 LLVIKDGDKKVGLYTVFVNPANVYQFVAGGQDQFMRIYDQRKIDENVNGVLKKFCPHHL 541

          610      620      630      640      650      660
NOV52      382 VNSESKANITCLVYSHDGTELLASYNDEDIYLFNSSHSDGAQYVKRYKGHRNNATVKGVN 441
gi|1363682| 382 VNSESKANITCLVYSHDGTELLASYNDEDIYLFNSSHSDGAQYVKRYKGHRNNATVKGVN 441
gi|7657148| 382 VNSESKANITCLVYSHDGTELLASYNDEDIYLFNSSHSDGAQYVKRYKGHRNNATVKGVN 441
gi|17485807| 565 VNCDFPTNITCVVYSHDGTELLASYNDEDIYLFNSSHSDGAQYVKRYKGHRNNITVKGVN 624
gi|17485821| 362 VYCDFPTNITCVVYSHDGTELLASYNDEDIYLFNSSLSDGAQYVKRYKGHRNNITVKGVN 421
gi|6679281| 542 LSSDYPAHITSLMYSYDGTETLLASYNDEDIYLFNSSDSDGAQYAKRYKGHRNNSTVKGVY 601

```


Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors. The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

In higher eukaryotes G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is also called a WD-40 repeat). Such a repetitive segment has been shown to exist in a number of other proteins, including G-beta-like peptides, yeast STE4, MS11, CDC4, CDC20, MAK11, PRP4, PWP1 and TUP1, slime-mould AAC3 and coronin, and Drosophila Groucho protein. The number of repeats within these proteins varies between 5 (PRP4, TUP1, and Groucho) and 8 (G-beta, STE4, MS11, AAC3, CDC4, PWP1, etc.). In G-beta and G-beta like proteins, the repeats span the entire length of the sequence, while in other proteins, they make up the N-terminal, the central or the C-terminal section.

The protein of this invention contains 7 WD-40 repeats. Although the function of this H326-like protein is not precisely known, it has potential importance in the intracellular transduction of signals, similarly to other WD-40 repeat-containing proteins.

The disclosed NOV52 nucleic acid of the invention encoding a H326-like protein includes the nucleic acid whose sequence is provided in Table 52A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 52A while still encoding a protein that maintains its H326-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1 percent of the bases may be so changed.

The disclosed NOV52 protein of the invention includes the H326-like protein whose sequence is provided in Table 52B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 52B while still encoding a protein that maintains its H326-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 36 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this H326 -like protein (NOV52) is a member of a "H326 family". Therefore, the NOV52 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV52 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in intracellular transduction of signals and any relevant diseases that may result from dysregulation of signal transduction, and/or other diseases and pathologies.

NOV52 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV52 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV52 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV53

A disclosed NOV53 nucleic acid of 1233 nucleotides (also referred to as CG56745-01) encoding a uracil phosphoribosyltransferase-like protein is shown in Table 53A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 142-144

and ending with a TAA codon at nucleotides 1069-1071. The start and stop codons are shown in bold in Table 53A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 53A. NOV53 nucleotide sequence (SEQ ID NO:195).

CTAGGGGTGAAAGGACAGCCAGGGTTAGATGTTCTGAGGAGGCGGGAGCAACCGAGAGAGCAGTGAGCATC
 TGTCTTTCTACCCGTTCCCTTTATCTTTAGTGTTCAGTAGCAGCGGGATAGCCCGGGGCCCCGGTGTATG
 GCCACGGAGTTACAGTGTCCGACTCCATGCCCTGTACACACAGCAAGTAACTCTGCCTCAACCCCAAGT
 CCCCAGCAGCTGCGACCTGGCGATCTGATCCTGGACCACGCAGGGGGAACAGAGCCTCCAGGGCCAAGGTG
 ATTCTCTCACGGGGTACGCCATTCTAGCCTGCCGGCCGAGCTGGACTCTGGGGCCTGCGGCGGCTCCAGC
 CTCAACTCAGAGGGCAACAGTGGTAGTGGTGACAGCTATGACGCACCAGCTGGCAACTCCTTCCTA
 GAGGACTGCCAACTCTCCCGCAGATCGGGGCGCAGCTTAAGCTGCTGCCTATGAATGATCAGATACGGGAG
 CTACAGACCATCATCCGGGACAAGACAGCCAGTAGAGGTGACTTCATGTTTTCTGCGGATCGTTTGATCAGA
 CTTGTTGTGGAAGAGGGATTGAATCAGCTGCCATATAAGAATGCATGGTGACCACTCCAACAGGGTACAAG
 TATGAAGGAGTGAAATTTGAGAAGGGAAATTGTGGGGTCAGCATAATGAGAAGCGGTGAGGCAATGGAACAA
 GGTTTACGAGACTGCTGTCGATCCATACGAATTGGAAGATCCTGATTACAGAGTGATGAGGAGACACAAAGA
 GCCAAAGTATATTATGCCAAATCCCCCAGACATTACCGGAGAAAAGTCCTTCTGATGTATCCAATTCTC
 AGCACTGGAAATACTGTAATTGAAGCTGTAAAGTTCTTATAGAACATGGAGTTCAACCCAGTGTATCATC
 CTACTCAGTCTGTTCTCCACTCCTCATGGTGCCAAATCAATCATTGAGGAGTTCCAGAGATCACAATTTTA
 ACTACTGAAGTTCATCTGTTGCACCTACACATTTTGGACAGAAATACTTTGGAACAGACTAAGTTATTTAA
 GTAAAATAATTGTCTTATGTAATATTACAATCATGTTTTGATTTCATTGTTTACTGATTCACTTGAGG
 GTGGCAGAGACAAATGTGTACAATGCTTTTAGTTTTGAAGTGGGTATATTTGAGGTATATCTCACTTA
 GTTATTTGT

In a search of public sequence databases, the NOV53 nucleic acid sequence, located on the X chromosome, has 312 of 544 bases (57%) identical to a gb:GENBANK-ID: YSCFUR1A|acc: M36485.1 mRNA from *Saccharomyces cerevisiae* (*S.cerevisiae* uracil phosphoribosyltransferase (FUR1) gene, complete cds).

The disclosed NOV53 polypeptide (SEQ ID NO:196) encoded by SEQ ID NO:195 has 309 amino acid residues and is presented in Table 53B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV53 has no signal peptide and is likely to be localized to the nucleus with a certainty of 0.3000. Alternatively, NOV53 may also localize to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 53B. Encoded NOV53 protein sequence (SEQ ID NO:196).

MATELQCPDSMPCHNQVNSASTPSPEQLRPGDLILDHAGGNRASRAKVILLTGyahssLPaelDSGACGGS
 SLNSEGNSGSDSSSYDAPAGNSFLEDCELSRQIGAQLKLLPMNDQIRELQTIIRDKTASRGDFMFSADRLI
 RLVVEEGLNQLPYKECMVTTPTGYKYEYGVKFEKGNCGVSIIRSGEAMEQGLRDCCRSIRIGKILIQSDEETQ
 RAKVYYAKFPPIYRRKVLMLPILSTGNTVIEAVKVLIEHGVQPSVILLSLFSTPHGAKSIIQEFPEITI
 LTTEVHPVAPTHFGQKYFGTD

A search of sequence databases reveals that the NOV53 amino acid sequence has 588 of 597 amino acid residues (98%) identical to, and 138 of 209 amino acid residues (66%) identical to, and 165 of 209 amino acid residues (78%) similar to, the 261 amino acid residue ptnr:SPTREMBL-ACC:Q9VRQ1 protein from *Drosophila melanogaster* (Fruit fly) (CG5537 Protein) ($E = 8.5e^{-72}$).

NOV53 is predicted to be expressed in at least Bone Marrow, Brain, Bladder, Eye, Cervix, Kidney, Liver, Lymph node, Prostate, Small Intestine, Umbilical Vein. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, and/or RACE sources.

NOV53 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 53C

Table 53C. BLAST results for NOV53					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17486179 ref XP_060041.1 (XM_060041)	similar to Unknown (protein for MGC:23937) (H. sapiens) [<i>Homo sapiens</i>]	309	309/309 (100%)	309/309 (100%)	e-172
gi 14388454 dbj BAB60766.1 (AB063019)	hypothetical protein [Macaca fascicularis]	309	300/309 (97%)	302/309 (97%)	e-167
gi 13874465 dbj BAB46861.1 (AB060829)	hypothetical protein [Macaca fascicularis]	309	299/309 (96%)	302/309 (96%)	e-166
gi 14388519 dbj BAB60785.1 (AB063065)	hypothetical protein [Macaca fascicularis]	309	298/309 (96%)	301/309 (96%)	e-166
gi 8217490 emb CAB92761.1 (AL137013)	bA311P8.3 (probable uracil phosphoribosyltra nferase) [<i>Homo sapiens</i>]	166	166/166 (100%)	166/166 (100%)	3e-94

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 53D. In the ClustalW alignment of the NOV53 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 53D. ClustalW Analysis of NOV53

- 1) Novel NOV53 (SEQ ID NO:196)
- 2) gi|17486179|ref|XP_060041.1| (XM_060041) similar to Unknown (protein for MGC:23937) (H. sapiens) [*Homo sapiens*] (SEQ ID NO:567)
- 3) gi|14388454|dbj|BAB60766.1| (AB063019) hypothetical protein [Macaca fascicularis] (SEQ ID NO:568)
- 4) gi|13874465|dbj|BAB46861.1| (AB060829) hypothetical protein [Macaca fascicularis] (SEQ ID NO:569)

5) gi|14388519|dbj|BAB60785.1| (AB063065) hypothetical protein [Macaca fascicularis] (SEQ ID NO:570)

6) gi|8217490|emb|CAB92761.1| (AL137013) bA311P8.3 (probable uracil phosphoribosyltransferase) [Homo sapiens] (SEQ ID NO:571)

5

10

15

20

25

30

35

40

45

50

55

NOV53

gi|17486179|

gi|14388454|

gi|13874465|

gi|14388519|

gi|8217490|

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

Table 53E lists the domain descriptions from DOMAIN analysis results against NOV53. This indicates that the NOV53 sequence has properties similar to those of other proteins known to contain this domain.

Table 53E Domain Analysis of NOV53

gnl|Pfam|pfam00156, Pribosyltran, Phosphoribosyl transferase domain. This family includes a range of diverse phosphoribosyl transferase enzymes. This family includes: Adenine phosphoribosyltransferase EC:2.4.2.7, Hypoxanthine-guanine-xanthine phosphoribosyltransferase, Hypoxanthine phosphoribosyltransferase EC:2.4.2.8, Ribose-phosphate pyrophosphokinase i EC:2.7.6.1, Amidophosphoribosyltransferase EC:2.4.2.14, Orotate phosphoribosyltransferase EC:2.4.2.10, Uracil phosphoribosyltransferase EC:2.4.2.9, Xanthine-guanine phosphoribosyltransferase EC:2.4.2.22. (SEQ ID NO:852)
CD-Length = 153 residues, 43.1% aligned
Score = 35.0 bits (79), Expect = 0.006

NOV53: 226 PPDIYRRKVLMLYPILSTGNTVIEAVKVLIEHGVQPSVILLSLFSTPHGAKSIIQEFPE 285
|+ ++||++ ++ || |+ |++| | + + +| + + + ||
Sbjct: 87 VGDVGGKRVLI VDDVIDTGGTIRAAAEELLKEAGAKVVGAVLVDRPEGGARERLDKGFPI 146
NOV53: 286 ITILTT 291
+++
Sbjct: 147 PSLIVL 152

10 The gene of invention is a novel uracil phosphoribosyltransferase (UPRT)-like gene. UPRT catalyzes the formation of uridine 5'-monophosphate in the pyrimidine salvage pathway from uracil and 5-phospho-alpha-D-ribose 1-diphosphate. The *Saccharomyces cerevisiae* FUR1 gene encodes UPRT (Kern et al., Gene 88:149-157(1990)). Mutations in the FUR1 gene have been correlated to resistance to 5-fluorouracil, a common chemotherapeutic agent (Kern et al., Curr Genet 1991 May;19(5):333-7).

The novel gene belongs to a family of phosphoribosyl transferases, as evidenced by the presence of a characteristic domain. It is anticipated that this gene plays a role in the pyrimidine salvage pathway and that it influences the growth or growth restriction of various tissues, and especially of tumor cells.

20 The disclosed NOV53 nucleic acid of the invention encoding a Uracil Phosphoribosyltransferase-like protein includes the nucleic acid whose sequence is provided in Table 53A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 53A while still encoding a protein that maintains its Uracil Phosphoribosyltransferase-like
25 activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications
30 include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least

in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 43 percent of the bases may be so changed.

5 The disclosed NOV53 protein of the invention includes the Uracil Phosphoribosyltransferase-like protein whose sequence is provided in Table 53B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 53B while still encoding a protein that maintains its Uracil Phosphoribosyltransferase-like activities and physiological functions, or a
10 functional fragment thereof. In the mutant or variant protein, up to about 4 percent of the residues may be so changed.

 The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

 The above disclosed information suggests that this Uracil Phosphoribosyltransferase-like protein (NOV53) is a member of a "Uracil Phosphoribosyltransferase family". Therefore, the NOV53 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to:
15 protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug
20 targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

 The NOV53 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease,
25 stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease, fertility disorders,
30 anemia, bleeding disorders, scleroderma, cystitis, incontinence, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, cirrhosis, inflammatory bowel disease, diverticular disease, lymphedema, cancer, trauma, tissue degeneration, bacterial/viral/parasitic infections, and/or other diseases and pathologies.

NOV53 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV53 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV53 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV54

NOV54 includes two novel protein phosphatase 2C -like proteins disclosed below. The disclosed sequences have been named NOV54a and NOV54b.

NOV54a

A disclosed NOV54a nucleic acid of 2185 nucleotides (also referred to as CG56773-01) encoding a protein phosphatase 2C-like protein is shown in Table 54A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 1402-1404. The start and stop codons are shown in bold in Table 54A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 54A. NOV54a nucleotide sequence (SEQ ID NO:197).

ATGTCCGCCGCGCTGGTTCGGCGCCGCTTCCTGCCTGGGGAGCCGCTCCCGCGCCGCGGCGCCCTGGGCCG
CATGCCAGCCCCGTGCCCTACCGACGGCCCCGCTTCCTTCGCGGCTCCAGCTCCAGCCCCGGGGCGGCCGAC
GCCTCGCGCCGCCAGACTCCCGGCCGTGCGCAGCCCCGCACGAGGACGCACGCTACCCCTGGAATGCAGGC
TACGCCGAGATTATCAATGCAGAGAAATCTGAATCAATGAGGATCAAGCCGCTGTGGGAAGCTGTGCATC
CGGAGATGTGAGTTTGGGGCTGAAGAAGAGTGGCTGACCCCTGTGCCAGAGGAGTTCCTGACAGGCCATTAC
TGGGCACTGTTCGATGGGCACGGCGTCTGCAGCAGCCATCTTGGCTGCCAACCCCTGCACTCCTGCTTG
CGCCGGCAGCTGGAGGCCGTGGTGAAGGCTTGGTGGCCACTCAGCCCCCATGCACCTCAATGGCCGCTGC
ATCTGCCCCAGTGACCCTCAGTTTGGGAGGAAAGGGCATCAGGGCAGAAGACTTGGTGATCGGGGCAATTG
GAGAGTGCCCTTCAGGAATGTGATGAGGTGATCGGGCGGGAGCTGGAGGCCTCAGGCCAGATGGGCGGCTGC
ACAGCCCTGGTGGCTGTGTCCCTGCAGGGAAAGCTGTACATGGCCAATGCTGGGGATAGCAGGGCCATCTTG
GTGCGGAGAGATGAGATACGGCCACTGAGCTTCGAGTTCACCCCAGAGACTGAGCGGCAGCGGATCCAGCAG
CTGGCCTTTGTCTATCCTGAGCTTCTGGCTGGTGAGTTCACCGACTGGAGTTCCTCGCGCGCTGAAGGGG
GATGACTTGGGACAGAAGGTTTTGTTCAGGGATCACACATGAGTGGCTGGAGCTACAAACGTGTGGAGAAA
TCGGATCTCAAGTACCCACTGATCCATGGACAGGGTAGGCAGGCTCGGTTACTAGGAACACTGGCTGTCTCC
CGGGCCCTGGGAGACCATCAGCTCAGAGTCTGGACACAAACATCCAGCTCAAGCCCTTCTTGCTCTCTGTG
CCACAGGTGACTGTGCTGGATGTGGACCACTGGAGCTACAGGAGGATGATGTGGTTGTCATGGCAACTGAT
GGACTCTGGGATGTACTGTCCAACGAGCAGGTGGCATGGCTGGTGCGGAGCTTCCTCCCTGGGAACCAAGAG
GACCCACACAGCTATCTGCAGGATGGTCTTCACAGGTTCTCAAAGCTGGCCAGATGCTGATACACAGCACA
CAGGGAAGGAAGACAGTCTCACAGAGGAAGGGCAGGTGCTCTACGATGACGCTCTGTGTTCGTGATTCCC
TTGCACAGTCAGGGCCAAGAGAGCAGTGACCACTGAGGATTTCAGACACTGTATCCCAAGACTGCTCTAGTGC
CCGGGTGTGGTCTGGGCATCCCTCCAGTGTGACCAAGAGCAAATCCTGCCTGCCCTATCCCTAGCCACCGCC
CAGTGTCTCTACTATCCACCTCAACACATCCATCTCAAGAGGAACATTTATACCAGGCAGTCAGAGCTGG
AAGTGTATGGAGAGCCAGCCACAGGTCTGCCTTTTGGCGTGATAACCTTCTCTGGCAGAGTGACTTTA
CAACTTAAGTAGGAAACCCATGTGAGGCTCTCAGACAGGATCTTGAACAGCCCAAAGTATCATTCTCAGAT
AGGGGCACCCAAGCTAAGGGTATTAGCCAAAGATGCCAGGATGGGTAGCTAGCCCATGTTTAGATCCAGGTC
TCCAATTCATGGTTATCAGGGCATGTGTTCAACAACCCCAAGTCCACGAGGTGGCTTTAGAAACCTTTT
GGGCAGCCTCATGTCTGCTAAACAGCCATCTTCAAGACAGCCCTGAAAAGAGACCAGTTCAGGTCTGGCC
CTGCTGTTCTTTGCTGGAGATGAGGAACAGGTGCTGGGGCTAAAGTTTGGGGTAGAGCACAAGGACAAGAG
GAACTCTTGGAGTTGGCTGGGTGAGAGGGCTCTCCATTGCTACCTGTAGTAGCTGCCTCTAACTGGTTG
CTTCTCCCTAGTTCCAGCCCTGCCTGGTCTGATGCCCCAACACTGCCCTTGCTTTGTTTCCCTGTACCT

CCCTATTATTAAATGTTTTCTACAG

In a search of public sequence databases, the NOV54a nucleic acid sequence, located on the p21.1 region of chromosome 3, has 592 of 928 bases (63%) identical to a gb:GENBANK-ID:AK023315|acc:AK023315.1 mRNA from *Homo sapiens* (cDNA

5 FLJ13253 fis, clone OVARC1000751) ($E = 4.5e^{-41}$).

A disclosed NOV54a polypeptide (SEQ ID NO:198) encoded by SEQ ID NO:197 has 467 amino acid residues and is presented in Table 54B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV54a has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.3941. Alternatively, 10 NOV54a may also localize to the nucleus with a certainty of 0.3000, to the lysosome (lumen) with a certainty of 0.1558, or to the mitochondrial matrix space with a certainty of 0.1000.

Table 54B. Encoded NOV54a protein sequence (SEQ ID NO:198).

MSAGWFRFRFLPGEPLPAPRPPGPHASVPYRRPRFLRGSSSSPGAADASRRPDSRPVRSARGRTLPPWNAG YAEIINA EKSEFNEDQAACGKLCIRRCFEFGAEEEWLTLCPPEFLTGHYWALFDGHHGPAAAILAANTLHSC RRQLEAVVEGLVATQPPMHLNGRCICPSDPQFVEEKGIRAE DLVIGALES AFQECDEVIGRELEASGQMGGC TALVAVSLQGKLYMANAGDSRAILVRRDEIRPLSFEFTPETERQRIQQALAFVYPELLAGEFTRLFFPRLKG DDLGGQKVLFRDHHMSGWSYKRVEKSDLKYPLIHGQGRQARLLGTLAVSRGLGDHQLRVLDNTIQLKPFLLSV PQVTVLDVDQLELQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQEDPHSYLQDGLHRFSKLAQMLIHST QGKEDSLTEEGQVSYYDDVSFVIPLHSQGQESSDH
--

A search of sequence databases reveals that the NOV54a amino acid sequence has 32 15 of 77 amino acid residues (41%) identical to, and 48 of 77 amino acid residues (62%) similar to, the 413 amino acid residue ptnr:SPTREMBL-ACC:Q9M3V1 protein from *Fagus sylvatica* (Beechnut) (Protein Phosphatase 2C (PP2C) (EC 3.1.3.16)) ($E = 9.2e^{-16}$).

NOV54a is predicted to be expressed in at least bone marrow, lymphoid tissue, tonsils, brain, colon, uterus, endometrium, placenta, mammary gland/breast, prostate, testis, foreskin, 20 heart, kidney, lung, spleen, peripheral blood, pituitary gland, retina, and pooled germ cell tumors. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

In addition, the sequence is predicted to be expressed in ovarian carcinoma because of 25 the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AK023315|acc:AK023315.1) a closely related *Homo sapiens* cDNA FLJ13253 fis, clone OVARC1000751 homolog.

NOV54b

A disclosed NOV54b nucleic acid of 1930 nucleotides (also referred to as CG56773- 02) encoding a protein phosphatase 2C-like protein is shown in Table 54C. An open reading

frame was identified beginning with a ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 1147-1149. The start and stop codons are shown in bold in Table 54C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 54C. NOV54b nucleotide sequence (SEQ ID NO:199).	
<p>ATGTCCGCCCGGCTGGTTCCGGCGCCGCTTCCTGCTGGGGAGCCGCTCCCGCGCCGCGGGCCGCTGGGCGG CATGCCAGCCCCGTGCCCTACCGACGGCCCCGCTTCCTTCGCGGCTCCAGCTCCAGCCCCGGGGCGGCCGAC GCCTCGCGCCGCCCAGACTCCCGGCCCGTGCAGCCCCGACAGGAGCGACGCTACCCCTGGAATGCAGGC TACGCCGAGATTATCAATGCAGAGAAATCTGAATTCAATGAGGATCAAGCCGCTGTGGGAAGCTGTGCATC CGGAGATGTGAGTTTGGGGCTGAAGAAGAGTGGCTGACCTGTGCCAGAGGAGGATGAGGTGATCGGGCGG GAGCTGGAGGCCTCAGGCCAGATGGGCGGCTGCACAGCCCTGGTGGCTGTGTCCCTGCAGGGAAAGCTGTAC ATGGCCAATGCTGGGGATAGCAGGGCCATCTTGGTGCGGAGAGATGAGATACGGCCACTGAGCTTCGAGTTT ACCCAGAGACTGAGCGGCAGCGGATCCAGCAGCTGGCCCTTGTCTATCCTGAGCTTCTGGCTGGTGAGTTC ACCCGACTGGAGTTCCCTCGGCGGCTGAAGGGGATGACTTGGGACAGAAGGTTTGTTCAGGGATCACCAC ATGAGTGGTGGAGCTACAAACGTGTGGAGAAATCGGATCTCAAGTACCCACTGATCCATGGACAGGGTAGG CAGGCTCGGTTACTAGGAACACTGGCTGTCTCCCGGGCTGGGAGACCATCAGCTCAGAGTCTTGGACACA AACATCCAGCTCAAGCCCTTCTTGCTCTCTGTGCCACAGGTGACTGTGCTGGATGTGGACCAGCTGGAGCTA CAGGAGGATGATGTGGTGTGCATGGCAACTGATGGACTCTGGGATGTACTGTCCAACGAGCAGGTGGCATGG CTGGTGCGGAGCTTCCTCCCTGGGAACCAAGAGGACCCACACAGCTATCTGCAGGATGGTCTTCACAGGTTT TCAAAGCTGGCCAGATGCTGATACACAGCACACAGGGAAGGAAGACAGTCTCAGAGAGGAAGGGCAGGTG TCCTACGATGACGTCTCTGTGTTCTGTGATTCCCTTGCACAGTCAGGGCCAAGAGAGCAGTGACCACTAGGA TTCAGACACTGTATCCAGAACTGCTCTAGTGCCCCGGGTGGTCTGGGCATCCCTCCAGTGTGACCAAGAG CAAATCCTGCCTGCCCTATCCCTAGCCACCGCCAGTGCTCTCACTATCCACCTCAACACACATCCATCTCA AGAGGAACATTTATACCAGGCAGTCAGAGCTGGAAGTGATGGAGAGCCAGCCACAGGTCTGCCTTTT CGGGTGATAACCTTCTCTGGCAGAGTGACTTTACAACCTTAAGTAAAGGATGAGGCTCCTCAGACAG GATCTTGAACAGCCCAAGTATCATTTCTCAGATAGGGGCACCCAAGCTAAGGGTATTAGCCAAAGATGCCAG GATGGGTAGCTAGCCCATGTTTAGATCCAGGTCTCCAATTCATGGTTATCAGGCATGTGTTCAACAACCC CAAAGTCCACGCAGGTGGCTTGTAGAAACCTTTGGGCAGCCTCATGTCTGCTAAAACAGCCATCTTCAAGAC AGCCCTGAAAAGAGACCAGTTACAGTCTGCTGCTGTTCTTTGTGGAGATGAGGAACAGGTGCTGGGG CTAAAGTTTGGGGTAGAGCACAAAGGACAAGAGGAACCTTGGAGTTGGGTGGGTGAGAGGGCTCTCCATTT GCTACCTGTAGTAGCCTGCCTCTTAAGTGGTTGCTTCTCCCTAGTTCCAGCCCTGGCTGTGATGCCCC AACACTGCCCTTGCTTTGTTTTCCCTGTACCTCCCTATTATTAATGTTTTCTACAG</p>	

- 5 In a search of public sequence databases, the NOV54b nucleic acid sequence, located on chromosome 3, has 446 of 660 bases (67%) identical to a gb:GENBANK-ID:BC011803|acc:BC011803.1 mRNA from *Homo sapiens* (*Homo sapiens*, Similar to RIKEN cDNA 2310008J22 gene, clone MGC:19531 IMAGE:4336762, mRNA, complete cds) (E = 2.0e⁻⁵⁶).
- 10 The disclosed NOV54b polypeptide (SEQ ID NO:200) encoded by SEQ ID NO:199 has 382 amino acid residues and is presented in Table 54D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV54b has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.4037. Alternatively, NOV54b may also localize to the nucleus with a certainty of 0.3000, to the
- 15 lysosome (lumen) with a certainty of 0.1000, or to the mitochondrial matrix space with a certainty of 0.1000.

Table 54D. Encoded NOV54b protein sequence (SEQ ID NO:200).	
MSAGWFRRRFLPGEPLPAPRPPGPHASVPVYRRPRFLRGSSSSPGAADASRRPDSRPVRSFARGRTLFPWNAG	

YAEIINA EKSEFNEDQAACGKLCIRRC EFGAEEEWLTLCP EEEDEVIGRELEASGQMGGCTALVAVSLQGKLY
MANAGDSRAILVRRDEIRPLSFEFTPE TERQRIQQ LAFVYPELLAGEFTRLEFP RRLKGDDLQKVLFRDHH
MSGWSYKRVEKSDLKYPLIHGQGRQARLLGTLAVSRGLGDHQLRVLD TN IQLKPFLLSV PQVTVLDVDQLEL
QEDDVVVMATDGLWDVLSNEQVAWLVR SFLPGNQEDPHSYLQDGLHRFSKLAQMLIHSTQGKEDSLTEEGQV
SYDDVSVFVIPLHSQGOESSDH

A search of sequence databases reveals that the NOV54b amino acid sequence has 231 of 270 amino acid residues (85%) identical to, and 244 of 270 amino acid residues (90%) similar to, the 453 amino acid residue ptnr:SPTREMBL-ACC:Q9CSD6 protein from *Mus musculus* (Mouse) (2810423O19RIK Protein) ($E = 1.6e^{-167}$).

NOV54b is predicted to be expressed in at least the following tissues: bone marrow, lymphoid tissue, tonsils, brain, colon, uterus, endometrium, placenta, mammary gland/breast, prostate, testis, foreskin, heart, kidney, lung, spleen, peripheral blood, pituitary gland, retina, and pooled germ cell tumors.

NOV54 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 54E

Table 54E. BLAST results for NOV54

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 12850332 dbj BAB 28679.1 (AK013149)	Protein phosphatase 2C containing protein-datasourc e: Pfam, source key: PF00481, evidence: ISS-puta tive[<i>Mus musculus</i>]	453	400/459 (87%)	423/459 (92%)	0.0
gi 16552416 dbj BAB 71302.1 (AK056894)	unnamed protein product [<i>Homo sapiens</i>]	270	251/252 (99%)	251/252 (99%)	e-143
gi 17462396 ref XP_ 059571.1 (XM_059571)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	247	247/255 (96%)	247/255 (96%)	e-137
gi 12856386 dbj BAB 30649.1 (AK017245)	Protein phosphatase 2C containing protein-datasourc e: Pfam, source key: PF00481, evidence: ISS-puta tive[<i>Mus musculus</i>]	254	222/261 (85%)	234/261 (89%)	e-121
gi 17455719 ref XP_ 051093.2 (XM_051093)	KIAA1157 protein [<i>Homo sapiens</i>]	514	213/480 (44%)	288/480 (59%)	6e-98

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 54F. In the ClustalW alignment of the NOV54 protein, as

well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 54F. ClustalW Analysis of NOV54

	1) Novel NOV54a (SEQ ID NO:198)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
--	---------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

	gi 17455719	178	VDILKNSAVLPPTCLGEEPENTPANSRTLTRAASLRGGVGAPGSPSTPPTRFFTEKKIPH	237				
		250	260	270	280	290	300	
5	NOV54a	185	EDLVIGALES	AFQECDEVIGRELEAS	GOMGGCTALVAVSLQGGKLYMANAGDSRAILVRRD		244	
	NOV54b	114	-----	DEVIGRELEAS	GOMGGCTALVAVSLQGGKLYMANAGDSRAILVRRD		159	
	gi 12850332	178	EDLVIGALENA	AFQECDDVIGRELEAS	GOMGGCTALVAVSLQGGKLYMANAGDSRAILVRRH		237	
	gi 16552416	24	EDLVIGALES	AFQECDEVIGRELEAS	GOMGGCTALVAVSLQGGKLYMANAGDSRAILVRRD		83	
	gi 17462396	1	-----	-----	MGGCTALVAVSLQGGKLYMANAGDSRAILVRRD		32	
10	gi 12856386	1	-----	-----	LEALGOVGGCTALVAVSLQGGKLYMANAGDSRAILVRRH		38	
	gi 17455719	238	ECLVIGALES	AFKEMDLQIERERSSYNIS	GGCTALVAVSLQGGKLYMANAGDSRAILVRRH		297	
		310	320	330	340	350	360	
15	NOV54a	245	EIRPLSF	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMSG		304	
	NOV54b	160	EIRPLSF	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMSG		219	
	gi 12850332	238	EIRQLS	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMRG		297	
	gi 16552416	84	EIRPLSF	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMSG		143	
	gi 17462396	33	EIRPLSF	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMSG		92	
20	gi 12856386	39	EIRQLS	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMRG		98	
	gi 17455719	298	EITPMS	FEFTPETERQRIQQLAFV	YPELLAGEFTRLEFPRLKGGDLGQKVLFRDHHMRG		357	
		370	380	390	400	410	420	
25	NOV54a	305	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	364
	NOV54b	220	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	279
	gi 12850332	298	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	357
	gi 16552416	144	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	203
	gi 17462396	93	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	152
30	gi 12856386	99	WSYKRVEK	SDLKYPLIHGQGRQARLLG	TLAVSRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	158
	gi 17455719	358	WAYKTIE	DELKPFLLSVCECKKARVMATIGVIRGLGDHQLRVLD	TNIQLKPFLLSV	PQVT	417	
		430	440	450	460	470	480	
35	NOV54a	365	VLDVDQLE	EQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		424
	NOV54b	280	VLDVDQLE	EQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		339
	gi 12850332	358	VLDVHOLA	VQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		410
	gi 16552416	204	VLDVDQLE	EQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		252
	gi 17462396	153	VLDVDQLE	EQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		204
40	gi 12856386	159	VLDVHOLA	VQEDDVVVMATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		211
	gi 17455719	418	TYDLSKYDHGSSDDVLILATDGLWDVLSNEQVAVLVRSLFPGNQED	PHSYLQDGLH	RFSEL		470	
		490	500	510	520			
45	NOV54a	425	AQMLIH	STQG---KEDSLTEEGQVSYDDVSVFVIPLHSQGOESSDH				467
	NOV54b	340	AQMLIH	STQG---KEDSLTEEGQVSYDDVSVFVIPLHSQGOESSDH				382
	gi 12850332	411	AKMLIH	NTQG---KDNATGEGQVSYDDVSVFVIPLHSQGOEGSGH				453
	gi 16552416	252	-----	CSCWG---PAWAVG---AS-----SKPK-----				270
	gi 17462396	205	AQMLIH	STQG---KEDSLTEEGQVSYDDVSVFVIPLHSQGOESSDH				247
50	gi 12856386	212	AKMLIH	NTQG---KDNATGEGQVSYDDVSVFVIPLHSQGOEGSGH				254
	gi 17455719	471	AQDLVM	RRARCVLKDRLGWRI	SNDRLGSGDDISVYVIPLHGNKLS---			514

Tables 54G-H list the domain descriptions from DOMAIN analysis results against NOV54. This indicates that the NOV54 sequence has properties similar to those of other proteins known to contain this domain.

Table 54G Domain Analysis of NOV54

gnl|Smart|smart00332, PP2Cc, Serine/threonine phosphatases, family 2C, catalytic domain; The protein architecture and deduced catalytic mechanism of PP2C phosphatases are similar to the PP1, PP2A, PP2B family of protein Ser/Thr phosphatases, with which PP2C shares no sequence similarity. (SEQ ID NO:853)
CD-Length = 260 residues, 73.5% aligned
Score = 114 bits (286), Expect = 9e-27

NOV54:	118	GHYWALFDGHGGPAAAILAANTLHSLRRQLEAVVEGLVATQPPMHLNGRCICPSDPQFV	177
		++ + + + +	
Sbjct:	40	GGFFGVFDGHGGSEAAKFLSKNLPEILAEELIKDKD-----	75
NOV54:	178	EEKGIRAE DLVIGALES AFQECDEVIGRELEASG-QMGGCTALVAVSLQGKLYMANAGDS	236
		++ + + +	
Sbjct:	76	-----EDEDVEDALRKAFRLTDEEILEELESLEDQRS GTTAVVALIRGNKLYVANVGDS	129
NOV54:	237	RAILVRRDEIRPLSFEFTPETERQRIQQALAFVYPELLAGEFTRLEFPRLKGDDLQKQVL	296
		+ + + + +	
Sbjct:	130	RAVLCRNGKAVQLTEDHKPSNEDER-----	154
NOV54:	297	FRDHHMSGWSYKRVEKSDLKYPLIHGQGRQARLLGTLAVSRGLGDHQLRVLDTNIQLKPF	356
		+ + ++ + + +	
Sbjct:	155	-----ERIREA-----GGFVSNGRVNGVLALSRLGDFFL-----KPY	187
NOV54:	357	LLSVPQVTVLDVDQLELQEDDVVVMATDGLWDVLSNEQVAWLVRSL	403
		+++ +++ + + + + + +	
Sbjct:	188	VIAEPDVTVELTEKDDFLI-----LASDGLWDVLSNQEVVDIVRKHL	230

Table 54H Domain Analysis of NOV54a

gnl|Pfam|pfam00481, PP2C, Protein phosphatase 2C. Protein phosphatase 2C is a Mn++ or Mg++ dependent protein S/threonine phosphatase. (SEQ ID NO:854)
CD-Length = 252 residues, 77.0% aligned
Score = 93.2 bits (230), Expect = 3e-20

NOV54:	119	HYWALFDGHGGPAAAILAANTLHSLRRQLEAVVEGLVATQPPMHLNGRCICPSDPQFVE	178
		++ + + + +	
Sbjct:	35	GFFAVFDGHGGSQAAYAGKHLETKLALR-----KSFPEL--	69
NOV54:	179	EKGIRAE DLVIGALES AFQEC-DEVIGRELEASGQMGGCTALVAVSLQGKLYMANAGDSR	237
		+ + + + + + +	
Sbjct:	70	-----DDLLENALKESFLESTDEELRSSAANTDLDSGSTAVVALIRGNKLYVANVGDSR	122
NOV54:	238	AILVRRDE-IRPLSFEFTP--ETERQRIQQALAFVYPELLAGEFTRLEFPRLKGDDLQKQ	294
		+ + + + +	
Sbjct:	123	AVLCRNGNAIKQLTEDHKPSNEDERRRIEGAG-----	154
NOV54:	295	VLFRDHHMSGWSYKRVEKSDLKYPLIHGQGRQARLLGTLAVSRGLGDHQLRVLDTNIQLK	354
		+ + + + +	
Sbjct:	155	-----GFVSRNGR--VNGVLAVSRAFGDFELK-----PGVL	183
NOV54:	355	PFLLSVPQVTVLDVDQLELQEDDVVVMATDGLWDVLSNEQVAWLVRSL	403
		+ + + + + + + +	
Sbjct:	184	QPVTAEPDVT----SHKITPSDEFLILASDGLWDVLSNQEVVDIVRSEL	228

Protein phosphorylation plays a key role in the regulation of cellular functions through the activation or inhibition of enzymes involved in various biochemical pathways. Kinases and

phosphatases that determine the phosphorylation state of an enzyme (and its activity) are frequently regulated through the action of hormones and growth factors (1). Four distinct subfamilies of serine/threonine protein phosphatases have been identified in mammals: PP1, PP2A, PP2B and PP2C (2). The PP2C subfamily contains structurally diverse protein phosphatases with a wide range of functions in cellular signal transduction; however, the exact physiological role of most PP2C enzymes is still unclear.

The protein described in this invention contains protein phosphatase 2C domains and is therefore likely to play a role in signal transduction and cellular proliferation. The protein is also homologous, but not identical, to the rat petrin protein, a PP2C subfamily member that has been shown to modulate neurite growth inhibition and may therefore be useful in the treatment of nerve damage resulting from traumatic injury, stroke or CNS degenerative disorders (3). The PP2C-like gene described in this invention is also expressed in the brain and may therefore have similar functions in the CNS. However, it is also expressed in a number of other tissues and based on its expression pattern may contribute to additional human diseases, such as cancer, inflammation/autoimmune diseases, and metabolic disorders. The PP2C-like gene maps to human chromosome 3p21.1.

Protein phosphatase 2C domain is found in protein phosphatase 2C, as well as other proteins e.g. adenylyl cyclase. Protein phosphatase 2C (PP2C) is one of the four major classes of mammalian serine/threonine specific protein phosphatases. PP2C is a monomeric enzyme of about 42 Kd that shows broad substrate specificity and is dependent on divalent cations (mainly manganese and magnesium) for its activity. Its exact physiological role is still unclear. Three isozymes are currently known in mammals: PP2C-alpha, -beta and -gamma. In yeast, there are at least four PP2C homologs: phosphatase PTC1 that has weak tyrosine phosphatase activity in addition to its activity on serines, phosphatases PTC2 and PTC3, and hypothetical protein YBR125c. Isozymes of PP2C are also known from *Arabidopsis thaliana* (ABI1, PPH1), *Caenorhabditis elegans* (FEM-2, F42G9.1, T23F11.1), *Leishmania chagasi* and *Paramecium tetraurelia*. In *Arabidopsis thaliana*, the kinase associated protein phosphatase (KAPP) is an enzyme that dephosphorylates the Ser/Thr receptor-like kinase RLK5 and which contains a C-terminal PP2C domain.

PP2C does not seem to be evolutionary related to the main family of serine/ threonine phosphatases: PP1, PP2A and PP2B. However, it is significantly similar to the catalytic subunit of pyruvate dehydrogenase phosphatase (PDPC), which catalyzes dephosphorylation and concomitant reactivation of the alpha subunit of the E1 component of the pyruvate

dehydrogenase complex. PDPC is a mitochondrial enzyme and, like PP2C, is magnesium-dependent.

The disclosed NOV54 nucleic acid of the invention encoding a Protein phosphatase 2C-like protein includes the nucleic acid whose sequence is provided in Table 54A, 54C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 54A or 54C while still encoding a protein that maintains its Protein phosphatase 2C-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37 percent of the bases may be so changed.

The disclosed NOV54 protein of the invention includes the Protein phosphatase 2C-like protein whose sequence is provided in Table 54B or 54D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 54B or 54D while still encoding a protein that maintains its Protein phosphatase 2C-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 56 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Protein phosphatase 2C-like protein (NOV54) is a member of a "Protein phosphatase 2C family". Therefore, the NOV54 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV54 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated cancer, trauma, bacterial and viral infections, in vitro and in vivo regeneration, fertility, endometriosis, hypogonadism, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD),

5 atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, anemia, Alzheimer's disease, stroke, hypercalcaemia, Parkinson's

10 disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, systemic lupus erythematosus, asthma, emphysema, allergy, ARDS, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Von Hippel-Lindau (VHL)

15 syndrome, endocrine dysfunctions, growth and reproductive disorders, tonsillitis, Hirschsprung's disease, Crohn's Disease, appendicitis, and/or other diseases and pathologies.

NOV54 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV54 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods

20 known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV54 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various

25 disorders.

NOV55

A disclosed NOV55 nucleic acid of 1500 nucleotides (also referred to as CG56806-01) encoding a Heparan Sulfate 6-Sulfotransferase 3-like protein is shown in Table 55A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 74-76 and

30 ending with a TGA codon at nucleotides 1490-1492. The start and stop codons are shown in bold in Table 55A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 55A. NOV55 nucleotide sequence (SEQ ID NO:201).

<p>CTTCCGAGCGGGCGCCCGTCCGCCCTGCCGCCCGCCCGCCCGCTTCGCCTGCCGGCCTGAGAGCGGGAC</p> <p>CATGGATGAAAGGTTCAACAAGTGGCTGCTGACGCCGGTGTCACTCTCCTCTTCGTGGTCATCATGTACCA</p> <p>GTACGTGTCCCCCTCCTGCACCACTCCTGCACCAACTTCGGGGAGCAGCCCCGCGGGGAGGCCCGCCCC</p> <p>GCCCCCGTCCCGGGTCCCGCCCGCGGGCTCAGGCGCCCGGAGAGTGGGAGCAGAGGAGGCCCGAGTT</p>
--

GCCCCCGCCCGCCCGGGGGCCCCCGAGGGACCTCGGGGGGCGCGCGCCGAGGAGGAGGACGAGGAGCC
 CGGAGACCCCGGGAGGGGGAGGAAGAGGAGGAGGAAGACGAGCCGACCCCGAGGCCCCGGAAAAACGGCTC
 CCTGCCCCGATTTCGTGCCGCGCTTCAACTTCAGCCTGAAGGACCTGACCCGCTTCGTGGATTTCACATCAA
 AGGGCGCGACGTGATCGTGTTCCTCCACATCCAGAAGACGGGGGACCACTTTCGGCCGGCACCTGGTGAA
 GAACATCCGGCTGGAGCAGCCTTGTAGCTGCAAAGCGGGTCAGAAGAAGTGACCTGCCACCGGCCTGGCAA
 GAAGGAGACGTGGCTCTTCTCCCGCTTCTCCACCGGCTGGAGCTGCGGGCTGCACGCCGACTGGACGGAGCT
 CACCAACTGCGTGCCCGCCATCATGGAGAAGAAGGACTGTCCCCGCAACCACAGCCACACCAGGAATTTCTA
 TTACATCACAAATGTTACGGGATCCAGTGTACGTTACCTGAGCGAGTGGAACATGTCCAGAGAGGGGCCAC
 TTGGAACCTCTCTCATATGTGTGATGGAAGAAGCCCCACCCAGATGAGCTGCCTACCTGCTACCCTGG
 GGATGACTGGTCTGGGGTCAGCTTGGCGGAGTTTATGGATTGCACCTACAACCTGGCTAACATCGCCAGGT
 GCGCATGCTGGCTGACCTCAGCCTGGTGGGCTGCTATAACTTGACTTTTCATGAACGAGAGTGAAAGAAACAC
 CATCTGTTCAGAGTGCAAAGAACAACCTGAAGAACATGGCCCTCTTTGGGCTCACTGAGTTCAGAGGAA
 GACACAGTTTCTCTTTCAGAGAACATTCAACCTCAAGTTCATCTCCCCCTTCACACAGTTCAACATCACGCG
 GGCTTCTAACGTGGAGATCAACGAGGGTGCCGCCAACGCATTGAGGATCTAACTTCTGGACATGCAGGT
 TTACGAGTATGCAAAGATCTCTTCCAGCAGCGCTACCACCACACCAAGCAGCTAGAGCACCAGAGGGACCG
 CCAGAAGCGCGGGAGGAGCGGAGGCTGCAGCGAGAGCACAGGGACCACAGTGGCCCCAAGAAGATGGGGC
 TGCAGAAGGGACTGTACCCGAGGACTACAACAGCCAGGTGGTGAGATGGTGACCTCCTGC

In a search of public sequence databases, the NOV55 nucleic acid sequence, located on chromosome 13, has 1329 of 1492 bases (89%) identical to a gb:GENBANK-
 ID:AB024567|acc:AB024567.1 mRNA from *Mus musculus* (mRNA for heparan sulfate 6-
 5 sulfotransferase 3, complete cds) ($E = 7.1e^{-263}$).

A disclosed NOV55 polypeptide (SEQ ID NO:202) encoded by SEQ ID NO:201 has 944 amino acid residues and is presented in Table 55B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV55 has a signal peptide and is likely to be localized to the plasma membrane with a certainty of 0.6850. Alternatively,
 10 NOV55 may also localize to the endoplasmic reticulum (membrane) with a certainty of 0.6400, to the Golgi body with a certainty of 0.3700, or to the microbody (peroxisome) with a certainty of 0.1269. The most likely cleavage site for NOV55 is between positions 28 and 29: VSP-SC.

Table 55B. Encoded NOV55 protein sequence (SEQ ID NO:202).

MDERFNKWLTPVLTLLFVIMYQYVSPSCTSSCTNFGEQPRAGEAGPPAVPGPARRAQAPPEWEQRRPQL
 PPPPRGPPEGPRGAAAPPEEEDDEPGDPREGEEEEDEPDPEAPENGSLPRFVPRFNFSCLKDLTRFVDFNFK
 GRDVIIVFLHIQKTGGTTTGRHLVKNIIRLEQPCSCKAGQKCTCHRPKKETWLFSTRFSTGWSCGLHADWTEL
 TNCVPAIMEKKDCPRNHSHTNRYIITMLRDPVSRYLSEWKHVQRGATWKTSLHMC DGRSPTPDELPTCYPG
 DDWSGVSLREFMDCTYNLANNRQVRMLADLSLVGCYNLTFMNERNTIILQSAKNNLKNMAFFGLTEFORK
 TQFLFERTFNLFKISPTQFNITRASNEINEGARQRIEDLNFLDMQLYEYAKDLFQQRYYHTKQLEHQDR
 QKRREERRLQREHRDHQWPKEGAAEGTVTEDYNSQVVRWMDERFNKWLTPVLTLLFVIMYQYVSPSCTS
 SCTNFGEQPRAGEAGPPAVPGPARRAQAPPEWEQRRPQLPPPRGPPEGPRGAAAPPEEEDDEPGDPREGEE
 EEEEEPDPEAPENGSLPRFVPRFNFSCLKDLTRFVDFNFKGRDVIIVFLHIQKTGGTTTGRHLVKNIIRLEQPC
 SCKAGQKCTCHRPKKETWLFSTRFSTGWSCGLHADWTELNCVPAIMEKKDCPRNHSHTNRYIITMLRDP
 VSRYLSEWKHVQRGATWKTSLHMC DGRSPTPDELPTCYPGDDWSGVSLREFMDCTYNLANNRQVRMLADLSL
 VGCYNLTFMNERNTIILQSAKNNLKNMAFFGLTEFORKTQFLFERTFNLFKISPTQFNITRASNEINE
 GARQRIEDLNFLDMQLYEYAKDLFQQRYYHTKQLEHQDRQKRREERRLQREHRDHQWPKEGAAEGTVTED
 YNSQVVRW

15

A search of sequence databases reveals that the NOV55 amino acid sequence has 447 of 472 amino acid residues (94%) identical to, and 458 of 472 amino acid residues (97%)

similar to, the 470 amino acid residue ptnr:SPTREMBL-ACC:Q9QYK4 protein from *Mus musculus* (Mouse) (Heparan Sulfate 6-Sulfotransferase 3) ($E = 7.5e^{-254}$).

NOV55 is predicted to be expressed in at least Right Cerebellum, Oviduct/Uterine Tube/Fallopian tube, Amygdala, and Kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

NOV55 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 55C

Table 55C. BLAST results for NOV55					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 7657192 ref NP_056635.1 (NM_015820)	heparan sulfate 6-O- sulfotransferase 3 [<i>Mus musculus</i>]	470	440/472 (93%)	452/472 (95%)	0.0
gi 16552186 dbj BAB71260.1 (AK056706)	unnamed protein product [<i>Homo sapiens</i>]	605	241/330 (73%)	284/330 (86%)	e-144
gi 14042611 dbj BAB55322.1 (AK027720)	unnamed protein product [<i>Homo sapiens</i>]	459	242/330 (73%)	285/330 (86%)	e-143
gi 7657190 ref NP_056634.1 (NM_015819)	heparan sulfate 6-O- sulfotransferase 2 [<i>Mus musculus</i>]	506	246/369 (66%)	290/369 (77%)	e-140
gi 12545389 ref NP_004798.2 (NM_004807)	heparan sulfate 6-O- sulfotransferase [<i>Homo sapiens</i>]	401	238/353 (67%)	286/353 (80%)	e-138

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 55D. In the ClustalW alignment of the NOV55 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 55D. ClustalW Analysis of NOV55

- 1) Novel NOV55a (SEQ ID NO:202)
- 2) gi|7657192|ref|NP_056635.1| (NM_015820) heparan sulfate 6-O-sulfotransferase 3 [*Mus musculus*] (SEQ ID NO:577)
- 3) gi|16552186|dbj|BAB71260.1| (AK056706) unnamed protein product [*Homo sapiens*] (SEQ ID NO:578)

- 4) gi|14042611|dbj|BAB55322.1| (AK027720) unnamed protein product [*Homo sapiens*]
(SEQ ID NO:579)
5) gi|7657190|ref|NP_056634.1| (NM_015819) heparan sulfate 6-O-sulfotransferase 2
[*Mus musculus*] (SEQ ID NO:580)
5 6) gi|12545389|ref|NP_004798.2| (NM_004807) heparan sulfate 6-O-sulfotransferase
[*Homo sapiens*] (SEQ ID NO:581)

			10	20	30	40	50	60	
10	NOV55	1	-----	-----	-----	-----	-----	-----	1
	gi 7657192	1	-----	-----	-----	-----	-----	-----	1
	gi 16552186	1	MALPACAVREFEP	PRQPERGAPVRT	TCPRRHSRVEAE	LAASRPGSVAAS	VRAGPPRGVSH		60
	gi 14042611	1	-----	-----	-----	-----	-----	-----	1
	gi 7657190	1	-----	-----	-----	-----	-----	-----	1
15	gi 12545389	1	-----	-----	-----	-----	-----	-----	1
			70	80	90	100	110	120	
20	NOV55	1	-----	-----	-----	-----	-----	-----	1
	gi 7657192	1	-----	-----	-----	-----	-----	-----	1
	gi 16552186	61	GFHTRPLLDKPR	KASSSLAGAACA	PLFALLSRGRRR	RMHVLRRRWDL	GLSLCRALLTR	GLA	120
	gi 14042611	1	-----	-----	-----	-----	-----	-----	1
	gi 7657190	1	-----	-----	-----	-----	-----	-----	1
25	gi 12545389	1	-----	-----	-----	-----	-----	-----	1
			130	140	150	160	170	180	
30	NOV55	1	-----	-----	-----	-----	-----	-----	34
	gi 7657192	1	-----	-----	-----	-----	-----	-----	34
	gi 16552186	121	ALGHSLKHVLGA	IFSKIFGPMASV	GNMDEKSNKLL	LALVMLFLFAV	IVLQYVCPGTE	CQ	179
	gi 14042611	1	-----	-----	-----	-----	-----	-----	33
	gi 7657190	1	-----	-----	-----	-----	-----	-----	33
	gi 12545389	1	-----	-----	-----	-----	-----	-----	31
35			190	200	210	220	230	240	
	NOV55	35	TNFGEQPRAGEA	GPPAVPGPARRA	QAAPPEEWEQR	RQPPEPRGPPE	GPRGAAPEEDE		94
	gi 7657192	35	TNFGEQPRAGEA	GPPAVPGPARRA	QAAPPEEWEQR	RQPPEPRGPPE	GPRGAAPEEDE		93
	gi 16552186	179	-----	-----	-----	-----	-----	-----	191
40	gi 14042611	33	-----	-----	-----	-----	-----	-----	45
	gi 7657190	33	-----	-----	-----	-----	-----	-----	45
	gi 12545389	31	-----	-----	-----	-----	-----	-----	50
45			250	260	270	280	290	300	
	NOV55	95	EPGDPRREGEE	EEEEDEPDPEA	PENGSLPRFVP	PRNFSLKDLTR	FVDENIKGRDV	IVFLHI	154
	gi 7657192	94	DPGDPEE-EE	EEEEDEPDPEA	PENGSLPRFVP	PRNFSLKDLTR	FVDENIKGRDV	IVFLHI	152
	gi 16552186	191	---GPYRSEDE	-----	-----	SSARFVPRYNF	TRGDLRLKVD	FDIKGDDIV	234
	gi 14042611	45	---DPYRSEDE	-----	-----	SSARFVPRYNF	TRGDLRLKVD	FDIKGDDIV	88
50	gi 7657190	45	---DPYRSEDE	-----	-----	SSARFVPRYNF	TRGDLRLKVD	FDIKGDDIV	88
	gi 12545389	50	---DP-----	-----	-----	HYEKKYYFPV	RELERSLRFD	MKGDDIV	84
55			310	320	330	340	350	360	
	NOV55	155	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	214
	gi 7657192	153	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	212
	gi 16552186	235	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	294
	gi 14042611	89	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	148
	gi 7657190	89	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	148
60	gi 12545389	85	QKTGGTTFGRH	LVRNIRLEQPC	SKAGQKKCTCH	RPGKRETWLF	SRSFGWSCGL	HADWT	144
65			370	380	390	400	410	420	
	NOV55	215	ELTNCVPAIME	KKDCPRNHSHT	-----	-----	-----	-----	236
	gi 7657192	213	ELTNCVPAIME	KKDCPRNHSHT	-----	-----	-----	-----	234
	gi 16552186	295	ELTSCVPSVVD	GKRDARLRPS	-----	-----	-----	-----	315
	gi 14042611	149	ELTSCVPSVVD	GKRDARLRPS	-----	-----	-----	-----	169
	gi 7657190	149	ELTSCVPSVVD	GKRDARLRPS	-----	-----	-----	-----	208
	gi 12545389	145	ELTNCVEGVLD	RRDSALR--TP	-----	-----	-----	-----	165

			430	440	450	460	470	480							
5	NOV55	236	-RNFYYITMLRDPVSRYLSEWKHVQ	GATWKTSLHMC	DGRSP	TPDELPTCY	PGDDWSGVS	295							
	gi 7657192	234	-RNFYYITMLRDPVSRYLSEWKHVQ	GATWKTSLHMC	DGRSP	TPDELPTCY	PGDDWSGVS	293							
	gi 16552186	315	-RNFHYITILRDPVSRYLSEWRHVQ	GATWKASLHV	CDGRPPTSEELP	SCYT	GDDWSGCP	374							
	gi 14042611	169	-RNFHYITILRDPVSRYLSEWRHVQ	GATWKASLHV	CDGRPPTSEELP	SCYT	GDDWSGCP	228							
	gi 7657190	209	GKRFHYITILRDPVSRYLSEWRHVQ	GATWKASLHV	CDGRPPTSEELP	SCYT	GDDWSGCP	268							
10	gi 12545389	165	-RKFFYYITILRDPVSRYLSEWRHVQ	GATWKTSLHMC	DGRTP	TPPEELPPCY	EGTDWSGCT	224							
			490	500	510	520	530	540							
15	NOV55	296	LREFMDCYNLANNRQVRMLADLS	LVGCYNLTFMNE	SERN	TILLOS	AKNNLKNMAFFGLT	355							
	gi 7657192	294	LREFMDCYNLANNRQVRMLADLS	LVGCYNLTFMNE	SERN	TILLOS	AKNNLKNMAFFGLT	353							
	gi 16552186	375	LKEFMDCPYNLANNRQVRMLS	DLTLVGCYNLSVMPE	KORKVLL	ESAKSNLKMAFFGLT	434								
	gi 14042611	229	LKEFMDCPYNLANNRQVRMLS	DLTLVGCYNLSVMPE	KORKVLL	ESAKSNLKMAFFGLT	288								
	gi 7657190	269	LKEFMDCPYNLANNRQVRMLS	DLTLVGCYNLSVMPE	KORKVLL	ESAKSNLKMAFFGLT	328								
20	gi 12545389	225	LKEFMDCPYNLANNRQVRMLADLS	LVGCYNLSFI	PEGKRAQL	LLES	AKNNLKNMAFFGLT	284							
			550	560	570	580	590	600							
25	NOV55	356	EFQRKTQYLFERTFNLKFI	SPFTQYNITRAS	NVEINEGAR	QRIEDLNFLDM	OLY	YAKDL	415						
	gi 7657192	354	EFQRKTQYLFERTFNLKFI	SPFTQYNITRAS	NVDINDGAR	QHIEELNFLDM	OLY	YAKDL	413						
	gi 16552186	435	EFQRKTQYLFERTFNMNFI	SPFTQYNITRAS	SVINEETQ	KRIEGLNFLDM	OLY	YAKDL	494						
	gi 14042611	289	EFQRKTQYLFERTFNMNFI	SPFTQYNITRAS	SVINEETQ	KRIEGLNFLDM	OLY	YAKDL	348						
	gi 7657190	329	EFQRKTQYLFERTFNMNFI	SPFTQYNITRAS	SVINEETQ	KRIEGLNFLDM	OLY	YAKDL	388						
30	gi 12545389	285	EFQRKTQYLFERTFNLKFI	RPFMQYNSTRAG	VEVDE	DTIRRIEELNFLDM	OLY	YAKDL	344						
			610	620	630	640	650	660							
35	NOV55	416	FOQRYHHTKOLEHORDR	OKRREER	RLQR	-----	-----	-----	443						
	gi 7657192	414	FOQRYHHTKOLEHORDR	OKRREER	RLQR	-----	-----	-----	441						
	gi 16552186	495	FLQRYQFMROKEHODARR	KRQEQCKFLKGRLLQ	THFQS	QGQGSQ	SN--	PNQNQ	SQNP	PNPN	552				
	gi 14042611	349	FLQRYQFMROKEHODARR	KRQEQCKFLKGRLLQ	THFQS	QGQGSQ	SN--	PNQNQ	SQNP	PNPN	406				
40	gi 7657190	389	FLQRYQFMROKEHODARR	KRQEQCKFLKGRFLQ	THFQS	QSQGSQ	SQS	SPGNL	SQNP	PNPN	448				
45	gi 12545389	345	FOQRYQYKROLERREOR	LRSREER	LLHR	-----	-----	-----	372						
			670	680	690	700	710								
45	NOV55	443	-----	EH	RD	-----	HQWPK	EDGAA	EGTV	-----	-----	472			
	gi 7657192	441	-----	EH	RA	-----	HRWPK	EDRAME	GTV	-----	-----	470			
	gi 16552186	553	ANQNLTQNLMQNLT	-----	QNLSQ	KENRES	PKNSG	KEQ	NDN	-	TSNGT	NDYIGS	VEK	WR	605
	gi 14042611	407	ANQNLTQNLMQNLT	-----	QSLSQ	KENRES	PKNSG	KEQ	NDN	-	TSNGT	NDYIGS	VEK	WR	459
	gi 7657190	449	PNQNLTQNLSHNLTP	SSNP	NSTORE	NRGSKQ	SGSGG	QGDS	SGT	SN	GT	NDYIGS	VEK	WR	506
45	gi 12545389	372	-----	AK	-----	EAL	P	REDA	DEP	GRVP	-----	-----	-----	-----	401

Heparan-sulfate 6-sulfotransferase (HS6ST) catalyzes the transfer of sulfate from 3'-phosphoadenosine 5'-phosphosulfate to position 6 of the N-sulfoglucosamine residue of heparan sulfate.

The disclosed NOV55 nucleic acid of the invention encoding a Heparan Sulfate 6-Sulfotransferase 3-like protein includes the nucleic acid whose sequence is provided in Table 55A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 55A while still encoding a protein that maintains its Heparan Sulfate 6-Sulfotransferase 3-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The

invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 11 percent of the bases may be so changed.

The disclosed NOV55 protein of the invention includes the Heparan Sulfate 6-Sulfotransferase 3-like protein whose sequence is provided in Table 55B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 55B while still encoding a protein that maintains its Heparan Sulfate 6-Sulfotransferase 3-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 34 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Heparan Sulfate 6-Sulfotransferase 3-like protein (NOV55) is a member of a "Heparan Sulfate 6-Sulfotransferase 3 family".

Therefore, the NOV55 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV55 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in Diabetes, Autoimmune disease, Renal artery stenosis, Interstitial nephritis, Glomerulonephritis, Polycystic kidney disease, Systemic lupus erythematosus, Renal tubular acidosis, IgA nephropathy, Hypercalcaemia, Lesch-Nyhan syndrome, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, Stroke, Tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, Cerebral palsy, Epilepsy, Lesch-Nyhan syndrome, Multiple sclerosis, Ataxia-telangiectasia, Leukodystrophies,

Behavioral disorders, Addiction, Anxiety, Pain, Neuroprotection, and/or other diseases and pathologies.

NOV55 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV55 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV55 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV56

NOV56 includes two novel N-Hydroxyarylamine Sulfotransferase-like proteins disclosed below. The disclosed sequences have been named NOV56a and NOV56b.

NOV56a

A disclosed NOV56a nucleic acid of 1223 nucleotides (also referred to as CG56816-01) encoding a N-Hydroxyarylamine Sulfotransferase-like protein is shown in Table 56A. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 2-4 and ending with a TGA codon at nucleotides 974-976. The start and stop codons are shown in bold in Table 56A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 56A. NOV56a nucleotide sequence (SEQ ID NO:203).

TATGGCGAAGATTGAGAAAAACGCTCCACGATGGAAAAAAGCCAGAAGTGTTTAACATCATGGAAGTAGA
TGGAGTCCCTACGTTGATATTATCAAAAGAATGGTGGGAAAAAGTATGTAATTTCCAAGCCAAGCCTGATGA
TCTTATTCATGCATCCATGTTGTACCTGACTTTGGGTAAGTTGCCAGAAGAAGATCATCAGGCTTGGCTTGG
AAATTACCCAAAGTCAGGTACAACATGGATGCATGAAATTTAGACATGATTCTAAATGATGGTGATGTGGA
GAAATGCAAAAGAGCCCAGACTCTAGATAGACACGCTTTCCCTTGAAGTGAATTTCCCATAAAGAAAAACC
AGATTTGGAGTTTCGTTCTTGAAATGTCCTCACCACAAGTATGAAAAACACATCTCCCTTCACATCTGATTCC
ACCATCTATCTGGAAAGAAAACTGCAAGATTGTCTATGTGACCAGAAATCCCAAGGATTGCC'TGGTGTCTTA
CTACCACTTTTACAGGATGGCTTCCTTTATGCCTGATCCTCAGAAGTATAGAGGAATTTTATGAGAAATTCAT
GTCCGAAAAAGTTGTTGGCGGGTCTCGGTTTGACCATATGAAAGGATGGTGGGCTGCAAAAGACATGCACCG
GATCCTCTACCTCTTCTACGAGGATATTAAAAAAATCCAAAACATGAGATCCACAAGGTGTTGGAATTCCTT
GGAGAAAACTTGGTCAGGTGATGTTATAAACAAGATTGTCCACCATACCTCATTTGATGTAATGAAGGATAA
TCCCATTGGCCAACCATACTGCGGTACCTGCTCACATATTCAATCACTCCATCTCAAAATTTATGAGGAAAGG
GATGCCCTGGAGACTGGAAGAACCACCTTTACTGTGGCTATGAATGAGAAGTGGATAAGCATTATGAAAAGAA
GATGGCAGGGTCCACACTGAACTTCTGCCTGGAGATCTGAGAGGAACAACAACAACTAGGTGACAGAGACT
ATGCCAACTATTTTCGCTTTTATTCTGTTGAGCAAGGAACTGTGACTGAATGTGGAGCTTATGAGCTTCAGT
CCATCTCCTATAGTGTGGCTAGTTTGCTATAATATTTAAACATGATTTAAATATCAACAAACAGTTACTC
CAGCAATAAAATAAGAGAATTAGAGACCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGG

In a search of public sequence databases, the NOV56a nucleic acid sequence, located on chromosome 2, has 633 of 921 bases (68%) identical to a gb:GENBANK-

ID:AF033653|acc:AF033653.1 mRNA from *Mus musculus* (phenol sulfotransferase mRNA, complete cds) (E = 7.3e⁻²⁷⁰).

A disclosed NOV56a polypeptide (SEQ ID NO:204) encoded by SEQ ID NO:203 has 324 amino acid residues and is presented in Table 56B using the one-letter amino acid code.

- 5 Signal P, Psort and/or Hydropathy results predict that NOV56a has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.7480. Alternatively, NOV56a may also localize to the mitochondrial membrane space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 56B. Encoded NOV56a protein sequence (SEQ ID NO:204).

MAKIEKNAPTMEKKPELPNIMEVDGVPTLLLSKEWWEKVCNFQAKPDDLIIHASMLYLTGLKLPPEEDHQAWLG NYPKSGTTWMHEILDMILNDGDVEKCKRAQTLDRHAFLELKFPHEKPDLEFVLEMSSPQLIKTHLP SHLIP PSIWKENCKIVYVTRNPKDCLVSYHFRMASFMPDPQNLEEFYEKFMMSGKVVGGSWFDHMKGWAAKDMHR ILYLFYEDIKKNPKHEIHKVLEFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMRKG MPGDWKNHFTVAMNENFDKHYEKKMAGSTLNFCEI

10

A search of sequence databases reveals that the NOV56a amino acid sequence has 155 of 254 amino acid residues (61%) identical to, and 196 of 254 amino acid residues (77%) similar to, the 304 amino acid residue ptmr:SWISSPROT-ACC:P50237 protein from *Rattus norvegicus* (Rat) (N-Hydroxyarylamine Sulfotransferase (EC 2.8.2.-) (HAST-I)) (E = 6.7e⁻⁹⁶).

15

NOV56b

- In the present invention, the target sequence identified previously, NOV56a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high
- 20
25
30

redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV56b. This is identical to the previously identified sequence (NOV56a).

A disclosed NOV56b nucleic acid of 1167 nucleotides (also referred to as CG56816-02) encoding a N-Hydroxyarylamine Sulfotransferase-like protein is shown in Table 56C. An open reading frame was identified beginning with a ATG initiation codon at nucleotides 33-35 and ending with a TGA codon at nucleotides 918-920. The start and stop codons are shown in bold in Table 56C, and the 5' and 3' untranslated regions, if any, are underlined.

Table 56C. NOV56b nucleotide sequence (SEQ ID NO:205).

TATGGCGAAGATTGAGAAAAACGCTCCCACGCATGGGAAAAGAAAGCCAGGAACTGTTTAAACATCATGGAAG
TAGATGGAGTCCCTACGTTGATATTATCAAAAGAAATGGTGGGAAAAGTATGTAATTTCCAAGCCAAGCCTG
ATGATCTTATTCTGGCAACTTACCCAAAGTCAGGTACAACATGGATGCATGAAATTTTAGACATGATTCTAA
ATGATGGTGATGTGGAGAAATGCAAAAGAGCCCAGACTCTAGATAGACACGCTTTCCTTGAAGTAAATTTTC
CCCATAAAGAAAAACCAGATTTGGAGTTCGTTCTTGAAATGTCCTCACCACAAGTATAAAAAACATCTCC
CTTCACATCTGATTCCACCATCTATCTGGAAAGAAAAGTCAAGATTGTCATGTGGCCAGAAATCCCAAGG
ATTGCCTGGTGTCTTACTACCACTTTACAGGATGGCTTCCTTTATGCCTGATCCTCAGAACTTAGAGGAAT
TTTATGAGAAATTCATGTCCGAAAAGTTGTTGGCGGGTCCCTGGTTTGACCATATGAAAGGATGGTGGGCTG
CAAAAGACATGCACCGGATCCTCTACCTCTTACGAGGATATTAAAAAAATCCAAAACATGAGATCCACA
AGGTGTTGGAATTCTTGGAGAAAAGTGGTCAGGTGATGTTATAAACAAGATTGTCCACCATACCTCATTTG
ATGTAATGAAGGATAATCCCATGGCCAACCATACTGCGGTACCTGCTCAGATATTCAATCACTCCATCTCAA
AATTTATGAGGAAAAGGATGCCTGGAGACTGGAAGAACCACCTTACTGTGGCTATGAATGAGAACTTTGATA
AGCATTATGAAAAGAAGATGGCAGGGTCCACACTGAACTTCTGCCTGGAGATCTGAGAGGAACAACAACA
CTAGGTGACAGAGACTATGCCAACTATTTCGCCCTTTATTCTGTTGAGCAAGGAAGTGTGACTGAATGTGGA
GCTTATGAGCTTCAGTCCATCTCTATAGTGTGGCTAGTTTGTATAATATTAACATGATTAAAAATATC
AACAAACCAAGTTACTCCAGCAATAAAATAAGAGAATTAGAGACCAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAGGC

In a search of public sequence databases, the NOV56b nucleic acid sequence, located on chromosome 2, has 649 of 919 bases (70%) identical to a gb:GENBANK-ID:AF033653|acc:AF033653.1 mRNA from *Mus musculus* (phenol sulfotransferase mRNA, complete cds) ($E = 3.6e^{-89}$).

The disclosed NOV56b polypeptide (SEQ ID NO:206) encoded by SEQ ID NO:205 has 295 amino acid residues and is presented in Table 56D using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV56b has no signal peptide and is likely to be localized to the microbody (peroxisome) with a certainty of 0.7480. Alternatively, NOV56b may also localize to the mitochondrial membrane space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 56D. Encoded NOV56b protein sequence (SEQ ID NO:206).

MGKESQELFNIMEVDGVPTLLLSKEWWEKVCNFKQAKPDDLILATYPKSGTTWMHEILDMILNDGDVEKCKRA
 QTLDRHAFLELKFPHEKPDLEFVLEMSSPQLIKTHLPShLI PPSIWKENCKIVYVARNPKDCLVSYHFFHR
 MASFMPPDPQNLEEFYEKFMGKVVGGSWFDHMKGWAAKDMHRILYLFYEDIKKNPKEIHKVLEFLEKTWS
 GDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMRKGMPPGDWKNHFTVAMNENFDKHYEKKMAGST
 LNFCLLEI

A search of sequence databases reveals that the NOV56b amino acid sequence has 173 of 283 amino acid residues (61%) identical to, and 220 of 283 amino acid residues (77%) similar to, the 304 amino acid residue ptrn:SWISSPROT-ACC:P50237 protein from *Rattus norvegicus* (Rat) (N-Hydroxyarylamine Sulfotransferase (EC 2.8.2.-) (HAST-I)) ($E = 4.0e^{-99}$).

NOV56b is predicted to be expressed in at least brain. .

NOV56a also has homology to the amino acid sequences shown in the BLASTP data listed in Table 56E

Table 56E. BLAST results for NOV56a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 17446341 ref XP_065757.1 (XM_065757)	similar to sulfotransferase, phenol preferring 2; Phenol sulfotransferase 1c1 (H. sapiens) [Homo sapiens]	304	299/324 (92%)	301/324 (92%)	e-166
gi 13929030 ref NP_113920.1 (NM_031732)	sulfotransferase, phenol preferring 2; Phenolsulfotransf erage 1c1 [Rattus norvegicus]	304	171/303 (56%)	218/303 (71%)	1e-94
gi 9055354 ref NP_061221.1 (NM_018751)	sulfotransferase, phenol preferring 2 [Mus musculus]	304	172/303 (56%)	217/303 (70%)	4e-94
gi 16304836 emb CAC95180.1 (AJ416889)	sulfotransferase 1C [Gallus gallus]	307	161/306 (52%)	214/306 (69%)	4e-88
gi 14731177 ref XP_010849.3 (XM_010849)	SULT1C sulfotransferase [Homo sapiens]	302	161/285 (56%)	201/285 (70%)	2e-85

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 56F. In the ClustalW alignment of the NOV56 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

Table 56F. ClustalW Analysis of NOV56

- 1) Novel NOV56a (SEQ ID NO:204)
- 2) Novel NOV56b (SEQ ID NO:206)
- 5) 3) gi|17446341|ref|XP_065757.1| (XM_065757) similar to sulfotransferase, phenol preferring 2; Phenol sulfotransferase 1c1 (*H. sapiens*) [*Homo sapiens*] (SEQ ID NO:582)
- 4) gi|13929030|ref|NP_113920.1| (NM_031732) sulfotransferase, phenol preferring 2; Phenolsulfotransferase 1c1 [*Rattus norvegicus*] (SEQ ID NO:583)
- 10) 5) gi|9055354|ref|NP_061221.1| (NM_018751) sulfotransferase, phenol preferring 2 [*Mus musculus*] (SEQ ID NO:584)
- 6) gi|16304836|emb|CAC95180.1| (AJ416889) sulfotransferase 1C [*Gallus gallus*] (SEQ ID NO:585)
- 15) 7) gi|14731177|ref|XP_010849.3| (XM_010849) SULT1C sulfotransferase [*Homo sapiens*] (SEQ ID NO:586)

		10	20	30	40	50	60	
20	NOV56a	MAKIEKNAPTMEKKPELFNIMEVDGVPTLILSKEWWEKVCNFOAKPDDLTHASMLYLTLG	60					
	NOV56b	MGKESQ-----ELFNIMEVDGVPTLILSKEWWEKVCNFOAKPDDLTLA	43					
	gi 17446341 ref	MAKIEKNAPTMEKKPELFNIMEVDGVPTLILSKEWWEKVCNFOAKPDDLTLA	52					
	gi 13929030 ref	MSLEKMKDLHLGEQDLQPETREVNGLMSKILMSDNWQKIWNFOAKPDDLTLA	52					
	gi 9055354 ref	MPLKLDLHLDEQNMQPETREVNGLMSKIMSENWQKIWNFOAKPDDLTLA	52					
	gi 16304836 emb	MALDKMENLSLEENMLRSEMGEVGGIPVTKPTCDIWDQVWNFKARPDDLVA	52					
25	gi 14731177 ref	MALHDMEDFTFDG-TKRLSVNYVKGILQPTDTCDIWDKIWNFOAKPDDLTLIS	51					
		70	80	90	100	110	120	
30	NOV56a	KLPEEDHQAWLGNYPKSGTTWMHEILDMLNDGDVEKCKRAQTLDRHAFLELKFP	117					
	NOV56b	-----TYPKSGTTWMHEILDMLNDGDVEKCKRAQTLDRHAFLELKFP	88					
	gi 17446341 ref	-----TYPKSGTTWMHEILDMLNDGDVEKCKRAQTLDRHAFLELKFP	97					
	gi 13929030 ref	-----TYAKAGTTWTQEIIVDMIONDGDVCKCORANTYDRHPFFLEWTLP	98					
	gi 9055354 ref	-----TYAKAGTTWTQEIIVDMIONDGDVCKCORANTYDRHPFFLEWTLP	98					
	gi 16304836 emb	-----TYAKAGTTWTQEIIVDMIOQNGDLEKCRRASTYKRHPFLEWYIPDSSPL	100					
35	gi 14731177 ref	-----TYPKAGTTWTQEIIVELTONEGDVERSKRAPTHORFPFLEMKIP	96					
		130	140	150	160	170	180	
40	NOV56a	EKPDLEFVLEMSSPOLIKTHLPVHMLPPSLWKENCKIYVYVTRNPKDCLVSYHHFHRMASF	177					
	NOV56b	EKPDLEFVLEMSSPOLIKTHLPVHMLPPSLWKENCKIYVYVARNPKDCLVSYHHFHRMASF	148					
	gi 17446341 ref	EKPDLEFVLEMSSPOLIKTHLPVHMLPPSLWKENCKIYVYVARNPKDCLVSYHHFHRMASF	157					
	gi 13929030 ref	N-SGLDLANKMPSPTLTKTHLPVHMLPPSFWKENSQIYVARNAKDCLVSYHHFHRNMK	157					
	gi 9055354 ref	N-SGLDLANKMPSPTLTKTHLPVQMLPPSFWKENSQIYVARNAKDCLVSYHHFHRNMK	157					
	gi 16304836 emb	GYSGLKLAELAMPSPRTMKTHLPVQLVPPSFWEONCKIYVARNAKDNLVSYHHFHRNMKV	160					
45	gi 14731177 ref	G-SGLEQAHAMPSPRILKTHLPVHMLPPSLLEKNCKIYVARNPKDNLVSYHHFHRMKA	155					
		190	200	210	220	230	240	
50	NOV56a	MPDPQNLLEEFYEKFMSSGKVVGGSWFDHMKGWAAAKDMHRILYLFYEDIKKNPKEIHKVL	237					
	NOV56b	MPDPQNLLEEFYEKFMSSGKVVGGSWFDHMKGWAAAKDMHRILYLFYEDIKKNPKEIHKVL	208					
	gi 17446341 ref	MPDPQNLLEEFYEKFMSSGKVVGGSWFDHVKGWAAAKDMHRILYLFYEDIKKNPKEIHKVL	217					
	gi 13929030 ref	LPDPGTLCEYIEQFKAGKVLGWSWYDHVKGWWDVQDKHRILYLFYEDMKEDPKREIKKIA	217					
	gi 9055354 ref	LPDPGTLCEYIETFKAGKVLGWSWYDHVKGWWDVQDKHRILYLFYEDMKEDPKREIKKIV	217					
	gi 16304836 emb	LPDPGTLCEETFEKFMNCEVLGWSWYDHVKGWWDVQDKHRILYLFYEDMKENPKREIKIM	220					
55	gi 14731177 ref	LPDPGTWEEYFETELAGKVCWGSWHVHVGWWEAKDKHRILYLFYEDMKKNPKHEIOKIA	215					
		250	260	270	280	290	300	
60	NOV56a	EFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMKRGMPGDWKNHF	297					
	NOV56b	EFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMKRGMPGDWKNHF	268					
	gi 17446341 ref	EFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMKRGMPGDWKNHF	277					
	gi 13929030 ref	KFLEKDISSEVLNKKIYHHTSFDVMKENPMANYTTLPSSIMDHSISPFMRKGMPPGDWKNYF	277					
	gi 9055354 ref	KFLEKDISSEVLNKKIYHHTSFDVMKENPMANYTTLPSSIMDHSISPFMRKGMPPGDWKNYF	277					
	gi 16304836 emb	KFLEKDLDEEVLNKKIYNTSFEIMKDNPMNYTKDFVGVMDSHVSFPMRKGSVGDWKNYF	280					
65	gi 14731177 ref	EFIGKKLLDKVLDKIVHYTSFDVMKDNPMANYSSIPAEIMDHSISPFMRKGAVDWKNHF	275					

5
 10
 NOV56a
 NOV56b
 gi|17446341|ref
 gi|13929030|ref
 gi|9055354|ref
 gi|16304836|emb
 gi|14731177|ref

310 320
 TVAMNENFDKHYEKKMAGSTLNFCLEI 324
 TVAMNENFDKHYEKKMAGSTLNFCLEI 295
 TVAMNENFDKHYEKKMAGSTLNFCLEI 304
 TVAQSEDFDEYRRKMAGSNITFRTEI 304
 TVAQSEDFDEYRRKMAGSNITFRTEI 304
 TVALNKKFDQDYKKKMADTSLVERMEI 307
 TVAQNERFDEYKKKMTDTRLTFHFQF 302

Tables 56G lists the domain descriptions from DOMAIN analysis results against NOV56. This indicates that the NOV56 sequence has properties similar to those of other proteins known to contain this domain.

Table 56G Domain Analysis of NOV56a

gnl|Pfam|pfam00685, Sulfotransfer, Sulfotransferase protein (SEQ ID NO:855)
 CD-Length = 269 residues, 99.6% aligned
 Score = 260 bits (665), Expect = 7e-71

20
 25
 30
 35
 NOV56: 25 GVPTLILSKEWKVCN-FOAKPDDLIHASMLYLTGKLPPEEDHQAWLGNYPKSGTTWMH 83
 | | | + | |||+|||+ + |||||+
 Sbjct: 2 GFWVDKFHLEGWRKIKNCFQARPDDV-----LIAGYPKSGTTWLQ 41
 NOV56: 84 EILDMILNDGDVEKCKRAQTLDRHAFLELKFPKHEKPDLEFVLEMSSPQLIKTHLP SHLI 143
 ||| + | || | | + + || + |||+||| || +
 Sbjct: 42 EILSLHPNVGDFEPPSPDPLLFNPNWLEYPKGEDWYETLKP--MPSSPRLIKTHLPLELL 99
 NOV56: 144 PPSIWKENCKIVYVTRNPKDCLVSYHFRMASFMPDPQN-LEEFYEKFMMSGKVVGGSWF 202
 | | ||+|| |||| ||||| | + ||| | ||+||| + |||
 Sbjct: 100 PKSFLSSKAKIIVLRNPKDVAVSYHFRSRSHKDLPADPGTFEEFLEAFNLGKVLYGSYF 159
 NOV56: 203 DHMKGWAAKDMHRILYLFYEDIKKNPKHEIHKVLEFLEKTWSGDVINKIVHHTSFDVMK 262
 ||+ ||| + ++|| |||+||| || | + ||| | + ++||| + ||| + |||
 Sbjct: 160 DHVLGWELRPEPQVLF LDYEDLKEDPAGEIKKIAEFLGLPLSEEELDKLLDHSSFFLMK 219
 NOV56: 263 DNPMAHNTAVPAHIFNHSISKFMKGMGPDWKNHFTVAMNENFDKHYEKK 312
 ||++| + + | ||||| + |||||+|| || ||| ++|
 Sbjct: 220 LNPLSNYETLCLGKSKGRKSPFMRKGLVGDWKNYFTPEQNEKFDKVIKEK 269

This protein carries out sulfation of phenols and bioactivation of n-hydroxyarylamines. It is responsible for the formation of N-hydroxy-2-acetylaminofluorene, a reactive metabolite which exhibits toxicity by binding to DNA, RNA and protein.

Hepatic sulfation of heterocyclic and non-heterocyclic arylamines was studied to assess enzymes responsible for their metabolisms. Both 2-amino-3-methylimidazo[4,5-f]quinoline (IQ)- and non-IQ-type (beta-carboline) heterocyclic amines were N-sulfated to form their sulfamates in cytosols of rat livers in the presence of 3'-phosphoadenosine-5'-phosphosulfate (PAPS). An arylsulfo-transferase, ST1A1, whose cDNA was isolated from a rat cDNA library, was expressed in COS-1 cells. The expressed enzyme catalyzed N-sulfation

of IQ, but not appreciably those of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), 2-amino-6-methyldipyrido[1,2-a:3',2'-d]imidazole (Glu-P-I), and 3-amino-1-methyl-5H-pyrido[4,3-b]indole (Trp-P-2). N-Sulfation of heterocyclic amines except IQ was higher in hepatic cytosols of female rats than of male rats. These results suggest the involvement of at least plural forms of sulfotransferase on the N-sulfation. In addition, N-sulfation of IQ was also observed in cytosol of a human liver, suggesting that N-sulfation is one of the metabolic pathways of heterocyclic amines in humans as well as rats. Hepatic sulfotransferase also catalyzes metabolic activation of N-hydroxy derivatives of carcinogenic arylamines. Using anti-HAST (hydroxylarylamine sulfotransferase) antibodies and ST1A1 cDNA as screening probes, several cDNA clones were isolated from the cDNA library. A new member of arylsulfotransferase, ST1C1, whose cDNA shows considerable sequence similarity to ST1A1 cDNA, was found to catalyze O-sulfation of N-hydroxy-2-acetylaminofluorence by the cDNA expression in COS-1 cells. From the close similarity of ontogenic profile and sex-specific expression of ST1C1 and HAST, ST1C1 cDNA was shown to encode a major sulfotransferase (HAST) mediating the metabolic activation of N-hydroxyarylamines in rat livers. In addition, properties of PAPS-dependent N-hydroxyarylamine activation and sulfotransferase in human livers are also discussed.

The disclosed NOV56 nucleic acid of the invention encoding a N-Hydroxyarylamine Sulfotransferase-like protein includes the nucleic acid whose sequence is provided in Table 56A, 56C or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 56A or 56C while still encoding a protein that maintains its N-Hydroxyarylamine Sulfotransferase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 48 percent of the bases may be so changed.

The disclosed NOV56 protein of the invention includes the N-Hydroxyarylamine Sulfotransferase-like protein whose sequence is provided in Table 56B or 56D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 56B or 56D while still encoding a protein that
5 maintains its N-Hydroxyarylamine Sulfotransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 32 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or (F_{ab})₂, that bind immunospecifically to any of the proteins of the invention.

10 The above disclosed information suggests that this N-Hydroxyarylamine Sulfotransferase-like protein (NOV56) is a member of a "N-Hydroxyarylamine Sulfotransferase family". Therefore, the NOV56 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this
15 invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

20 The NOV56 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in metabolic diseases and disorders, and/or other diseases and pathologies.

NOV56 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV56 substances for use in
25 therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV56 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in
30 understanding of pathology of the disease and development of new drug targets for various disorders.

NOV57

A disclosed NOV57 nucleic acid of 953 nucleotides (also referred to as CG56829-01) encoding a Testis Specific Serine Kinase-3-like protein is shown in Table 57A. An open

reading frame was identified beginning with a ATG initiation codon at nucleotides 50-52 and ending with a TGA codon at nucleotides 854-856. The start and stop codons are shown in bold in Table 57A, and the 5' and 3' untranslated regions, if any, are underlined.

Table 57A. NOV57 nucleotide sequence (SEQ ID NO:207).

CAGAGGCAGCATGAGCTGAGAGGGTGATAGGAAGGCGCGCTAGACAGC**ATGGAGGACTTCTGCTCTCCAA**
TGGGTACCAGCTGGGCAAGACCATGGGGAAGGGACCTACTCAAAAGTCAAGAAGCATTTC**CAAAAAACA**
CCAAAGAAAAGTGGCAATTAAGTTATAGACAAGATGGGAACTCTCAGAGTTATCCAGAGATTCTCTCC
TCGGGAGCTCCAAATCGTCCGTACCCTGGACCACAAGAACATCATCCAGGTGTATGAGATGCTGGAGTCTGC
CGACGGGAAAATCTGCCTGGTGTAGGAGCTCGCTGAGGGAGGGGATGTCTTTGACTGCGTGCTGAATGGGGG
GCCACTGCCTGAAAGCCGGGCCAAGGCCCTCTCCGTCAGATGGTTGAGGCCATCCGCTACTGCCATGGCTG
TGGTGTGGCCACCGGGACCTCAAATGTGAGAACGCCTTGTTCAGGGCTTCAACCTGAAGCTGACTGACTT
TGGCTTTGCCAAGGTGTGCCCCAAGTCACACCGGGAGCTGAGCCAGACCTTCTGCGGCAGTACAGCCTATGC
TGCCCCCGAGGTGCTGCAGGGCATTCCCCCACGATAGCAAAAAAGGTGATGTCTGGAGCATGGGTGTGGTCTCT
GTATGTCATGCTCTGTGCCAGCCTACCTTTTGACGACACAGACATCCCCAAGATGCTGTGGCAGCAGCAGAA
GGGGGTGTCTTCCCACTCATCTGAGCATCTCGGCCGATTGCCAGGACCTGCTCAAGAGGCTCCTGGAACC
CGATATGATCTCCGGCTTCAATTGAAGAAGTTAGTTGGCATCCATGGCTAGCAAGCACTTGATAAAAGCA
ATGGCAAGTGCTCTCCAATAAAGTAGGGGGAGAAAGCAAAACCCAAACCCGCTTCTAAAAATGGTGATATAT
ATTTTACGCTTAAGTT

In a search of public sequence databases, the NOV57 nucleic acid sequence, located on chromosome 1, has 831 of 912 bases (91%) identical to a gb:GENBANK-ID:AF201734|acc:AF201734.1 mRNA from *Mus musculus* (testis specific serine kinase-3 (Tssk-3) mRNA, complete cds) ($E = 2.0e^{-165}$).

A disclosed NOV57 polypeptide (SEQ ID NO:208) encoded by SEQ ID NO:207 has 268 amino acid residues and is presented in Table 57B using the one-letter amino acid code. Signal P, Psort and/or Hydropathy results predict that NOV57 has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. Alternatively, NOV57 may also localize to the microbody (peroxisome) with a certainty of 0.1821, to the mitochondrial matrix space with a certainty of 0.1000, or to the lysosome (lumen) with a certainty of 0.1000.

Table 57B. Encoded NOV57 protein sequence (SEQ ID NO:208).

MEDFLLSNGYQLGKTIGEGTYSKVKEAFSKKHQRKVAIKVIDKMGTSSSEFIQRFPLPRELQIVRTLQDHKNIIQ
 VYEMLESADGKICLVMELEAGGDVFDCLNNGGPLPESRAKALFRQMVEAIRYCHGCGVAHRDLKCENALLQG
 FNLKLTDPGFAKVLPSHRELSQTFCGSTAYAAPEVLQGI PHDSKKGDVWSMGVVLYVMLCASLPFDDTDIP
 KMLWQQQKGVSFPTHLSISADCQDLLKRLLEPDMILRPSIEEVSWHPWLAST

A search of sequence databases reveals that the NOV57 amino acid sequence has 240 of 268 amino acid residues (89%) identical to, and 245 of 268 amino acid residues (91%) similar to, the 266 amino acid residue ptnr:SPTREMBL-ACC:Q9JL98 protein from *Mus musculus* (Mouse) (Testis Specific Serine Kinase-3) ($E = 6.2e^{-124}$).

NOV57 is predicted to be expressed in at least lung, testis, B-cell, brain, head and neck. Expression information was derived from the tissue sources of the sequences that were

included in the derivation of the sequence of CuraGen Acc. No. CG56829-01. The sequence is predicted to be expressed in testis because of the expression pattern of (GENBANK-ID: gb:GENBANK-ID:AF201734|acc:AF201734.1) a closely related *Mus musculus* testis specific serine kinase-3 (Tssk-3) mRNA, complete cds homolog.

5 NOV57 also has homology to the amino acid sequences shown in the BLASTP data listed in Table 57C

Table 57C. BLAST results for NOV57					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi 16418343 ref NP_443073.1 (NM_052841)	testis-specific serine/threonine kinase 22C [<i>Homo sapiens</i>]	268	265/268 (98%)	265/268 (98%)	e-157
gi 12860201 dbj BAB31876.1 (AK019840)	data source:MGD, source key:MGI:1929914, evidence:ISS-putative-serine/threonine kinase 22C (spermiogenesis associated) [<i>Mus musculus</i>]	268	259/268 (96%)	261/268 (96%)	e-153
gi 10946880 ref NP_067454.1 (NM_021479)	serine/threonine kinase 22C (spermiogenesis associated); testis specific serine kinase-3 (Tssk-3) [<i>Mus musculus</i>]	266	240/273 (87%)	245/273 (88%)	e-136
gi 16507245 ref NP_443732.1 (NM_053006)	serine/threonine kinase 22B (spermiogenesis associated); testis specific serine threonine kinase 2 [<i>Homo sapiens</i>]	358	127/266 (47%)	192/266 (71%)	1e-70
gi 14776972 ref XP_033051.1 (XM_033051)	serine/threonine kinase 22B (spermiogenesis associated) [<i>Homo sapiens</i>]	358	127/266 (47%)	192/266 (71%)	1e-70

The homology between these and other sequences is shown graphically in the ClustalW analysis shown in Table 57D. In the ClustalW alignment of the NOV57 protein, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be altered to a much broader extent without altering protein structure or function.

10

15

Table 57D. ClustalW Analysis of NOV57

- 1) Novel NOV57 (SEQ ID NO:208)
- 2) gi|16418343|ref|NP_443073.1| (NM_052841) testis-specific serine/threonine kinase 22C [*Homo sapiens*] (SEQ ID NO:587)
- 3) gi|12860201|dbj|BAB31876.1| (AK019840) data source:MGD, source key:MGI:1929914, evidence:ISS-putative-serine/threonine kinase 22C (spermiogenesis associated) [*Mus musculus*] (SEQ ID NO:588)
- 4) gi|10946880|ref|NP_067454.1| (NM_021479) serine/threonine kinase 22C (spermiogenesis associated); testis specific serine kinase-3 (Tssk-3) [*Mus musculus*] (SEQ ID NO:589)
- 5) gi|16507245|ref|NP_443732.1| (NM_053006) serine/threonine kinase 22B (spermiogenesis associated); testis specific serine threonine kinase 2 [*Homo sapiens*] (SEQ ID NO:590)
- 6) gi|14776972|ref|XP_033051.1| (XM_033051) serine/threonine kinase 22B (spermiogenesis associated) [*Homo sapiens*] (SEQ ID NO:591)

			10	20	30	40	50	60	
20	NOV57		MED--	FLLSNGYQLGKTIGEGTYSKVKEAFSKKHQRKVAIKVIDKMGTSSEFIQRFPLPRE					58
	gi 16418343 ref		MED--	FLLSNGYQLGKTIGEGTYSKVKEAFSKKHQRKVAIKVIDKMGCPPEEFIQRFPLPRE					58
	gi 12860201 dbj		MED--	FLLSNGYQLGKTIGEGTYSKVKEAFSKKHQRKVAIKIIDKMGCPPEEFIQRFPLPRE					58
	gi 10946880 ref		MED--	FLLSNGYQLGKTIGEGTYSKVKEAFSKKHQRKVAIKIIDKMGCPPEEFIQRFPLPRE					58
	gi 16507245 ref		MDATVLRKKGYIVGINLCKGSAKVKSAYSERLKFNVAVKIIDRKKTPDTDFVERFLPRE						60
25	gi 14776972 ref		MDATVLRKKGYIVGINLCKGSAKVKSAYSERLKFNVAVKIIDRKKTPDTDFVERFLPRE						60
			70	80	90	100	110	120	
30	NOV57		LQIVRTLDDHKNI IQVYEMLESADGKI CLVMELAEAGGDVFDVCLNNGGPLPESRAKALFROM						118
	gi 16418343 ref		LQIVRTLDDHKNI IQVYEMLESADGKI CLVMELAEAGGDVFDVCLNNGGPLPESRAKALFROM						118
	gi 12860201 dbj		LQIVRTLDDHKNI IQVYEMLESADGKI CLVMELAEAGGDVFDVCLNNGGPLPESRAKALFROM						118
	gi 10946880 ref		LQIVRTLDDHKNI IQVYEMLESADGKI CLVMELAEAGGDVFDVCLNNGGPLPES--ARPSSARW						117
	gi 16507245 ref		MDILATVNHGSI IKTYEIFETS DGRYIYIMELGVQGDILFFIKCOGALHEDVARKMFROI						120
35	gi 14776972 ref		MDILATVNHGSI IKTYEIFETS DGRYIYIMELGVQGDILFFIKCOGALHEDVARKMFROI						120
			130	140	150	160	170	180	
40	NOV57		VEATRYCHGCGVAHRDLKCNALLQG--FNLKLTDFGFAK--VLPKS--HRELSQTFCGSTA						174
	gi 16418343 ref		VEATRYCHGCGVAHRDLKCNALLQG--FNLKLTDFGFAK--VLPKS--HRELSQTFCGSTA						174
	gi 12860201 dbj		VEATRYCHGCGVAHRDLKCNALLQG--FNLKLTDFGFAK--VLPKS--HRELSQTFCGSTA						174
	gi 10946880 ref		LRLFAIAMAV--RGHRDLKCNALLQG--FNLKLTDFGFAK--VLPKS--HRELSQTFCGSTA						172
	gi 16507245 ref		SSAVKYCHOLDIVHRDLKCNALLDKDFNIIKLSDFGFSKRCLRDSNGRIILSKTFCGSAA						180
	gi 14776972 ref		SSAVKYCHOLDIVHRDLKCNALLDKDFNIIKLSDFGFSKRCLRDSNGRIILSKTFCGSAA						180
45			190	200	210	220	230	240	
	NOV57		YAAPEVLQGI PHDSKKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQQKG--VSFPPTHLS						233
	gi 16418343 ref		YAAPEVLQGI PHDSKKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQQKG--VSFPPTHLS						233
	gi 12860201 dbj		YAAPEVLQGI PHDSKKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQQKG--VSFPPTHLS						233
	gi 10946880 ref		YAAPEVLQGI PHDSKKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQQKG--VSFPPTHLS						231
	gi 16507245 ref		YAAPEVLQSIPIYQPKVYDIWSLGVILYIMVCGSMPIYDDSDIRKMLRIQKEHRVDFPERSKN						240
50	gi 14776972 ref		YAAPEVLQSIPIYQPKVYDIWSLGVILYIMVCGSMPIYDDSDIRKMLRIQKEHRVDFPERSKN						240
			250	260	270	280	290	300	
55	NOV57		ISADCDLLKRLLEPDMILRPSIEEVSWHPWLAST						268
	gi 16418343 ref		ISADCDLLKRLLEPDMILRPSIEEVSWHPWLAST						268
	gi 12860201 dbj		ISTECQDLLKRLLEPDMILRPSIEEVSWHPWLAST						268
	gi 10946880 ref		ISTECQDLLKRLLEPDMILRPSIEEVSWHPWLAST						266
	gi 16507245 ref		LTCECKDLTYRMLOPDVSQRLLHIDEILSHSWLQPPKPKPTSSASFKEGEGKYRAECKLD						300
60	gi 14776972 ref		LTCECKDLTYRMLOPDVSQRLLHIDEILSHSWLQPPKPKATSSASFKEGEGKYRAECKLD						300
			310	320	330	340	350		
65	NOV57		-----	-----	-----	-----	-----		268
	gi 16418343 ref		-----	-----	-----	-----	-----		268

gi 12860201 dbj	-----	268
gi 10946880 ref	-----	266
gi 16507245 ref	TKTGLRPDHRPDHKLGAQTQHRLLVVPENENRMEDRLAETSRAKDHHISGAEVGKAST	358
gi 14776972 ref	TKTGLRPDHRPDHKLGAQTQHRLLVVPENENRMEDRLAETSRAKDHHISGAEVGKAST	358

Tables 57E-G lists the domain descriptions from DOMAIN analysis results against NOV57. This indicates that the NOV57 sequence has properties similar to those of other proteins known to contain this domain.

Table 57E Domain Analysis of NOV57

gnl|Smart|smart00220, S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. (SEQ ID NO:856)
 CD-Length = 256 residues, 100.0% aligned
 Score = 258 bits (659), Expect = 3e-70

NOV57:	10	YQLGKTIGEGTYSKVKEAFSKKHQRKVAIKVIDKMGTSSEFIQRFLPRELQIVRTL DHKN	69
		+ + + + + + + + ++ ++	
Sbjct:	1	YELLEVLGKGAFGKVYLARDKKTGKLVAIKVIKKEKLKKKKRERIL-REIKILKKLDHPN	59
NOV57:	70	IIQVYEMLESADGKICLVMELEAGGDVDFCVLNGGGLPESRAKALFRQMVEAIRYCHGCG	129
		+++ ++ + + + + ++ +	
Sbjct:	60	IVKLYDVFEEDDKLYLVMEYCEGGDLFDLLKKRGRLSEDEARFYARQILSALEYLHSQG	118
NOV57:	130	VAHRDLKCENALL-QGFNLKLTDFGFAKVLPKSHRELSQTFCGSTAYAAPEVLQGI PHDS	188
		+ ++ + +	
Sbjct:	119	IIHRDLKPENILLSDGHVKLADFGFLAKQL-DSGGTLLTTFVGTPEYMAPEVLLGKGYG-	176
NOV57:	189	KKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQ---QKGVSFPTHLSISADCQDLLKRL	245
		+ + + + + +++ + + + +	
Sbjct:	177	KAVDIWSLGVILYELLTGKPPFPFGDDQLLALFKKIGKPPPPFPPEWKISPEAKDLIKKL	236
NOV57:	246	LEPDMILRPSIEEVSWHPWL	265
		+ +	
Sbjct:	237	LVKDPEKRLTAEAELEHPFF	256

Table 57F Domain Analysis of NOV57

gnl|Pfam|pfam00069, pkinase, Protein kinase domain (SEQ ID NO:857)
 CD-Length = 256 residues, 100.0% aligned
 Score = 234 bits (596), Expect = 6e-63

NOV57:	10	YQLGKTIGEGTYSKVKEAFSKKHQRKVAIKVIDKMGTSSEFIQRFLPRELQIVRTL DHKN	69
		+ + + + + ++ + ++ ++ +	
Sbjct:	1	YELGEKLGSGAFGKVYKKGKHDGTGEIVAIIKILKKRSLSEK--KKRFLREIQILRRLSHPN	58
NOV57:	70	IIQVYEMLESADGKICLVMELEAGGDVDFCVL-NGGGLPESRAKALFRQMVEAIRYCHGC	128
		+++ + + + + + ++ +	
Sbjct:	59	IVRLLGVFEE-DDHLYLVMEYMEGGDLFDYLRRNGLLLLSEKEAKKIALQILRGLEYLHSR	117
NOV57:	129	GVAHRDLKCENALL-QGFNLKLTDFGFAKVLPKSHRELSQTFCGSTAYAAPEVLQGI PHD	187
		+ + + + + + +	
Sbjct:	118	GIVHRDLKPENILLDENGTVKIDFGLARKLESSSYEKLTTFVGTPEYMAPEVLEG-RGY	176
NOV57:	188	SKKGDVWSMGVVLYVMLCASLPFDDTDIPKMLWQQQKGVSFPTHLSISADCQDLLKRL	245
		+ + + + + ++ ++ + + + + +	

Sbjct: 177 SSKVDVWSLGVILYELLTGKLPFPGIDPLEELFRIKERPRLRLPLPPNCSEELKDLIKKC 236
 NOV57: 246 LEPDMILRPSIEEVSWHPWL 265
 | | ||+ +|+ |||
 5 Sbjct: 237 LNKDPEKRPTAKEILNHPWF 256

Table 57G Domain Analysis of NOV57

gnl|Smart|smart00219, TyrKc, Tyrosine kinase, catalytic domain;
 Phosphotransferases. Tyrosine-specific kinase subfamily (SEQ ID
 NO:858)
 CD-Length = 258 residues, 97.7% aligned
 Score = 115 bits (289), Expect = 2e-27

NOV57: 12 LGKTIGEGTYSKVKEAF--SKKHQRKVAIKVIDKMGTSSEFIQRFLPRELQIVRTLCHK 68
 ||| +||| + +| + + +||| + | | + | + || || ++| |||
 10 Sbjct: 3 LGKKLGEAFGEVYKGLTKGKGVEVEVAVKTL-KEDASEQQIEEFL-REARLMRKLDHP 60
 NOV57: 69 NIIQVYEMLESADGKICLVMELAEAGDVFCVLNGGP--LPESRAKALFRQMVEAIRYCH 126
 ||+++ + + + + +||| |||| + | + | | + | + + |
 15 Sbjct: 61 NIVKLLGVC-TEEPLMIVMEYMEGGDLLDYLRKNRPKELSLDLLSFALQIARGMEYLE 119
 NOV57: 127 GCGVAHRDLKCNALL-QGFNLKLTDFGFAK-VLPKSHRELSQTFCGSTAYAAPEVLQGI 184
 |||| | | + + +| + ||| + + + ++ + ||| | +
 Sbjct: 120 SKNFVHRDLAARNCLVGENKTVKIADFGGLARDLYDDDYRKKKSPRLPIRWMAPESLKDG 179
 20 NOV57: 185 PHDSKKGDVWSMGVVLVYVML-CASLPFDDTDIPKMLWQQQKGVSFPTHLSISADCQDLLK 243
 | | |||| | +| + + | + ++| +|| | + + || +
 Sbjct: 180 KFTS-KSDVWSFGVLLWEIFTLGESPYPGMSNEEVLEYLKKGYRLPQPPNCPDEIYDLML 238
 25 NOV57: 244 RLLEPDMILRPSIEEV 259
 + | ||+ | +
 Sbjct: 239 QCWAEDPEDRPTFSEL 254

Eukaryotic protein kinases (1) are enzymes that belong to a very extensive family of
 30 proteins which share a conserved catalytic core common with both serine/threonine and
 tyrosine protein kinases. Protein phosphorylation is a fundamental process for the regulation of
 cellular functions. The coordinated action of both protein kinases and phosphatases controls
 the levels of phosphorylation and, hence, the activity of specific target proteins. One of the
 predominant roles of protein phosphorylation is in signal transduction, where extracellular
 35 signals are amplified and propagated by a cascade of protein phosphorylation and
 dephosphorylation events. Two of the best characterized signal transduction pathways involve
 the cAMP-dependent protein kinase and protein kinase C (PKC). Each pathway uses a
 different second-messenger molecule to activate the protein kinase, which, in turn,
 phosphorylates specific target molecules. Extensive comparisons of kinase sequences defined
 40 a common catalytic domain, ranging from 250 to 300 amino acids. This domain contains key
 amino acids conserved between kinases and are thought to play an essential role in catalysis.
 In the N-terminal extremity of the catalytic domain there is a glycine-rich stretch of residues in
 the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the

central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme (2).

Protein kinases and phosphatases regulate cell-cycle progression, transcription, translation, protein sorting and cell adhesion events that are critical to the inflammatory process. Two of the best-characterized immunosuppressants, cyclosporin and rapamycin, are also effective anti-inflammatory drugs. They act directly on protein phosphorylation and, as such, validate the concept that small-molecule modulators of phosphorylation cascades possess anti-inflammatory properties (3).

Some examples of the role of serine/threonine protein kinases that are important in cell proliferation and disease include AKT, RAF1 and PIM1. Dudek et al. (4) demonstrated that AKT is important for the survival of cerebellar neurons. Thus, the 'orphan' kinase moved center stage as a crucial regulator of life and death decisions emanating from the cell membrane. Holland et al. (5) transferred, in a tissue-specific manner, genes encoding activated forms of Ras and Akt to astrocytes and neural progenitors in mice. These authors found that although neither activated Ras nor Akt alone was sufficient to induce glioblastoma multiforme (GBM) formation, the combination of activated Ras and Akt induced high-grade gliomas with the histologic features of human GBMs. These tumors appeared to arise after gene transfer to neural progenitors, but not after transfer to differentiated astrocytes. Increased activity of Ras is found in many human GBMs and Akt activity is increased in most of these tumors, implying that combined activation of these 2 pathways accurately models the biology of this disease (5).

Another disease that involves yet another serine/threonine kinase is Peutz-Jeghers syndrome (PJS), an autosomal dominant disorder characterized by melanocytic macules of the lips, buccal mucosa, and digits, multiple gastrointestinal hamartomatous polyps, and an increased risk of various neoplasms. Jenne et al. (6) identified and characterized the serine/threonine kinase STK11 and identified mutations in PJS patients. All 5 germline mutations were predicted to disrupt the function of the kinase domain. They concluded that germline mutations in STK11, probably in conjunction with acquired genetic defects of the second allele in somatic cells according to the Knudson model, caused the manifestations of PJS. These authors commented that PJS was the first cancer susceptibility syndrome identified that is due to inactivating mutations in a protein kinase and found mutations in the STK11 gene in 11 of 12 unrelated families with PJS. Ten of the 11 were truncating mutations. All were heterozygous in the germline. Su et al. (7) found that of 53 PJS patients with cancer reported to that time, 6 (11%) were diagnosed with pancreatic adenocarcinoma. Su et al. (7)

presented evidence that the STK11 gene plays a role in the development of both sporadic and familial (PJS) pancreatic and biliary cancers. They found that in sporadic cancers, the STK11 gene was somatically mutated in 5% of pancreatic cancers and in at least 6% of biliary cancers examined. In the patient with pancreatic cancer associated with PJS, there was inheritance of a mutated copy of the STK11 gene and somatic loss of the remaining wildtype allele.

The novel human serine/threonine protein kinase of the invention contains a protein kinase domain. Therefore it is anticipated that this novel protein has a role in the regulation of essentially all cellular functions and could be a potentially important target for drugs. Such drugs may have important therapeutic applications, such as treating numerous inflammatory diseases.

The disclosed NOV57 nucleic acid of the invention encoding a Testis Specific Serine Kinase-3-like protein includes the nucleic acid whose sequence is provided in Table 57A or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 57A while still encoding a protein that maintains its Testis Specific Serine Kinase-3-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 9 percent of the bases may be so changed.

The disclosed NOV57 protein of the invention includes the Testis Specific Serine Kinase-3-like protein whose sequence is provided in Table 57B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 57B while still encoding a protein that maintains its Testis Specific Serine Kinase-3-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 53 percent of the residues may be so changed.

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

The above disclosed information suggests that this Testis Specific Serine Kinase-3-like protein (NOV57) is a member of a "Testis Specific Serine Kinase-3 family". Therefore, the NOV57 nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The NOV57 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in metabolic diseases and disorders, and/or other diseases and pathologies.

NOV57 nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOV57 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below. The disclosed NOV57 protein has multiple hydrophilic regions, each of which can be used as an immunogen. These novel proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

NOV58

NOV58 includes two Gap Junction Beta-5-like proteins, designated herein as NOV58a and NOV58b. Gap junctions are conduits that allow the direct cell-to-cell passage of small cytoplasmic molecules, including ions, metabolic intermediates, and second messengers, and that thereby mediate intercellular metabolic and electrical communication. Gap junction channels consist of connexin protein subunits, which are encoded by a multigene family.

NOV58a

The disclosed NOV58a (alternatively referred to herein as CG56315-01) includes the 728 nucleotide sequence (SEQ ID NO:209) shown in Table 58A. A NOV58a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 28-30 and ends with a stop codon at nucleotides 697-699. The disclosed NOV58a maps to human chromosome 6.

Table 58A. NOV58a Nucleotide Sequence (SEQ ID NO:209)

AGTCTTGCCTTCTTTGAGCCTAAGTCATGAGTTGGATGTTCTCAGAGATCTCCTGAGT
GGAGTAAATAAATACTCCACTGGGACTGGATGGATTGGCTGGCTGTCGTGTTGTCTTC
CGTTTGTGGTCTACATGGTGGCAGCAGAGCACGTGTGAAAGATGAGCAGAAAGAGTTT
GAGTGCAACAGTAGACAGCCCGGTTGCAAAAATGTGTGTTTGTGACTTCTTCCCAT
TCCCAAGTCAGACTTTGGGCCCTTACAAGTATAATGGTCTCCACACCTTCACTTCTGGTG
GTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGCACAGAAAGAACTCTATGTC
AGCCCAGGTACAATGGATGGGGCCTATGGTACGCTTATCTTATCAGCCTCATTGTTAAA
ACTGGTTTGAATTTGGCTTCCTTGTGTTTATTTTATAAGCTATATGATGGCTTAGTGTT
CCCTACCTTATAAAGTGTGATTGAAGCCTTGTCCCAACTGTGGACTGCTTCATCTCC
AAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGATT
GTGTTGAATTTCAATGAAGTGAAGTGTGTTTGGTCTCAAGTGCTTATTAAGTGCTGTCTC
CAAAAATATTTAAAAAACCTCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGT
TGAATGTG

A NOV58a polypeptide (SEQ ID NO:210) encoded by SEQ ID NO:209 is 223 amino acids in length and is presented using the one-letter amino acid code in Table 58B. The Psort profile for NOV58a predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. In alternative embodiments, a NOV58a polypeptide is located to the mitochondrial inner membrane with a certainty of 0.4358, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the Golgi with a certainty of 0.4000. The Signal P predicts a likely cleavage site for a NOV58a peptide is between positions 40 and 41, *i.e.*, at the dash in the sequence VAA-EH.

Table 58B. NOV58a Polypeptide Sequence (SEQ ID NO:210)

MSWMFLRDLLSGVNKYSTGTGWIWLVVVFVRLLVYMVAEHVWKDEQKEFECNSRQPGC
KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRKKLYVSPGTMGGGL
WYAYLISLIVKTGFELVFLVLYDGFSPYLIKCDLKPCPNTVDCFISKPTKTIIFI
LFLVITSCLCIVLNFIELSFLVLKCFIKCLQKYLKKPQVLSV

NOV58b

The disclosed NOV58b (alternatively referred to herein as CG56315-02) includes the 727 nucleotide sequence (SEQ ID NO:211) shown in Table 58C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 27-29 and ends with a stop codon at nucleotides 696-698. The disclosed NOV58b maps to human chromosome 10.

Table 58C. NOV58b Nucleotide Sequence (SEQ ID NO:211)

AGTCTTGCTTCTTTTGGAGCCTAAGTCATGAGTTGGATGTTCTCAGAGATCTCCTGAGTG

GAGTAAATAAATACTCCACTGGGATTGGATGGATTGGCTGGCTGTCGTGTTGTCTTCC
 GTTTGCTGGTCTACATGGTGGCAGCAGACGTGTGGAAAGATGAGCAGAAAGAGTTTG
 AGTGCAACAGTAGACAGCCCGTTGCAAAAATGTGTGTTTGTGACTTCTTCCCCATTT
 CCCAAGTCAGACTTTGGGCCCTTACAACGTATAATGGTCTCCACACCTTCACCTCTGGTGG
 TTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGCACAGAAAGAACTCTATGTCA
 GCCCAGGTACAATGGATGGGGCCTATGGTACGCTTATCTTATCAGCCTCATGTGTTAAAA
 CTGGTTTTGAAATTGGCTTCCTTGT'TTTATTTTATAAGCTATATGATGGCTTTAGTGTTC
 CCTACCTTATAAAGTGTGATTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCCA
 AACCCTGAGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATTG
 TGTTGAATTTTCATGAACTGAGTTTTTTGGTTCTCAAGTGCTTTATTAAGTGCTGCTCCTC
 AAAAATATTTAAAAAACCTCAAGTCCTCAGTGTGTGAGTGCCACAGCCTCAGATATGTT
 GAATGTG

A NOV58b polypeptide (SEQ ID NO:212) encoded by SEQ ID NO:211 is 628 amino acids in length and is presented using the one-letter amino acid code in Table 58D. The Psort profile for NOV58b predicts that this sequence has a signal peptide and is likely to be
 5 localized at the plasma membrane with a certainty of 0.6000. The Signal P predicts a likely cleavage site for a NOV58b peptide is between positions 40 and 41, *i.e.*, at the dash in the sequence VAA-EH.

Table 58D. NOV58b Polypeptide Sequence (SEQ ID NO:212)

MSWMFLRDLLSGVNKYSTGIGWIWLAVVFVFRLLVYMVA AEHVWKDEQKEFECNSRQPGC
 KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRHKLYVSPGTMDDGL
 WYAYLISLIVKTGFEIGFLVLFYKLYDGFSPYLIKCDLKPCPNTVDCFISKPTKTIPI
 LFLVITSCLCIVLNFIELSFLVLKCFIKCCLQKYLKKPQVLSV

10 A BLAST analysis of NOV58 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV58 had high homology to other proteins as shown in Table 58E.

Table 58E. BLASTX results from PatP database for NOV58

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAV32079 Human gap junction protein beta-4	689	1.2e-67
patp:AAV36145 Human secreted protein #17 - <i>Homo sapiens</i>	689	1.2e-67
patp:AAV36192 Human secreted protein #64 - <i>Homo sapiens</i>	689	1.2e-67
patp:AAV70457 Human membrane channel protein-7 (MECHP-7)	666	3.3e-65
patp:AAG74001 Human colon cancer antigen protein	657	3.0e-64

15 In a search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 240 of 336 bases (71%) identical to a gb:GENBANK-ID:RATCXN311A|acc:M76533.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* connexin (CXN-311) gene). The full amino acid sequence of the protein of the invention was

found to have 125 of 220 amino acid residues (56%) identical to, and 170 of 220 amino acid residues (77%) similar to, the 271 amino acid residue ptrn:SWISSNEW-ACC:Q02739 protein from *Mus musculus* (Mouse) (GAP JUNCTION BETA-5 PROTEIN (CONNEXIN 31.1) (CX31.1)). NOV58 also has homology to the other proteins shown in the BLASTP data in Table 58F.

Table 58F. NOV58 BLASTP results

This BLASTP data is displayed graphically in the ClustalW in Table 58G. A multiple sequence alignment is given, with the NOV58a and b protein being shown on line 1 and 2 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 58F.

Table 58G. ClustalW Alignment of NOV58

		70	80	90	100	110	120	
NOV58a	61	KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRHKK	-----	LYVSP	113			
NOV58b	61	KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRHKK	-----	LYVSP	113			
gi 15990851	61	KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRHKK	-----	LYVSP	113			
gi 10835079	61	SNVCFDFFPVSQVRLWALQLILVTCPSLLVVMHVAYREVQEKRRHREAHGENSGRLYLNP	-----	LYVSP	120			
gi 6753996	61	TNVCYDEFFPVSQVRLWALQLILVTCPSLLVVMHVAYRKAREKKYQEKIG--EGYLYPNP	-----	LYVSP	118			
gi 15029850	61	TNVCYDEFFPVSQVRLWALQLILVTCPSLLVVMHVAYRKAREKKYQEKIG--EGYLYPNP	-----	LYVSP	118			
gi 4009522	61	SNVCFDFFPVSQVRLWALQLILVTCPSLLVVMHVAYREVQEKRRHREAHGENSGRLYLNP	-----	LYVSP	120			
		130	140	150	160	170	180	
NOV58a	114	GTMDGGLWYAYLISLIVKVTGFEEIGFLVLFYKLYDGFVSVPYLIKCDLKPCPNTVDCFISKP	-----	LYVSP	173			
NOV58b	114	GTMDGGLWYAYLISLIVKVTGFEEIGFLVLFYKLYDGFVSVPYLIKCDLKPCPNTVDCFISKP	-----	LYVSP	173			
gi 15990851	114	GTMDGGLWYAYLISLIVKVTGFEEIGFLVLFYKLYDGFVSVPYLIKCDLKPCPNTVDCFISKP	-----	LYVSP	173			
gi 10835079	121	GKKRGGLWWTYYVCSLIFKASVDIAFLYVHFHSFYPKYILPPVVKCHADPCPNIVDCFISKP	-----	LYVSP	180			
gi 6753996	119	GKKRGGLWWTYYVCSLIFKATIDIFLYLHFHFYPRYTLPSMVKCHAEPCPNTVDCFIAPK	-----	LYVSP	178			
gi 15029850	119	GKKRGGLWWTYYVCSLIFKATIDIFLYLHFHFYPRYTLPSMVKCHAEPCPNTVDCFIAPK	-----	LYVSP	178			
gi 4009522	121	GKKRGGLWWTYYVCSLIFKASVDIAFLYVHFHSFYPKYILPPVVKCHADPCPNIVDCFISKP	-----	LYVSP	180			
		190	200	210	220	230	240	
NOV58a	174	TEKTIFILELIVITSCICIVLNFIELSELVLKCFIKCLOKY	-----	LKK	217			
NOV58b	174	TEKTIFILELIVITSCICIVLNFIELSELVLKCFIKCLOKY	-----	LKK	217			
gi 15990851	174	TEKTIFILELIVITSCICIVLNFIELSELVLKCFIKCLOKY	-----	LKK	217			
gi 10835079	181	SEKNIFTLEMVATAATCILLNLVELIYLVSKRCHBCLAARKAQAMCTGHHPHGTTSSCKO	-----	LKK	240			
gi 6753996	179	SEKNIFTLEMVVTVAVICILLNLVELIYLVKRCSECAQLRRPPTAHAKNDPNWANSPSKE	-----	LKK	238			
gi 15029850	179	SEKNIFTLEMVVTVAVICILLNLVELIYLVKRCSECAQLRRPPTAHAKNDPNWANSPSKE	-----	LKK	238			
gi 4009522	181	SEKNIFTLEMVATAATCILLNLVELIYLVSKRCHBCLAARKAQAMCTGHHPHGTTSSCKO	-----	LKK	240			
		250	260	270				
NOV58a	218	PQVLSV	-----	223				
NOV58b	218	PQVLSV	-----	223				
gi 15990851	218	PQVLSV	-----	223				
gi 10835079	241	DDLLSGDLIFLGSDSHPPLLPDRPRDHVKKTIL	-----	273				
gi 6753996	239	KDFLSGDLIFLGSDAHPPLLDRPRAHVKKTIL	-----	271				
gi 15029850	239	KDFLSGDLIFLGSDAHPPLLDRPRAHVKKTIL	-----	271				
gi 4009522	241	DDLLSGDLIFLGSDSHPPLLPDRPRDHVKKTIL	-----	273				

The presence of identifiable domains in NOV58 was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (<http://www.ebi.ac.uk/interpro>). DOMAIN results for NOV58 as disclosed in Table 58H, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 58H fully conserved single residues are indicated by the sign (!) and “strong” semi-conserved residues are indicated by the sign (+). The “strong” group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Table 58H lists the domain description from DOMAIN analysis results against NOV58. This indicates that the NOV58 sequence has properties similar to those of other proteins known to contain this domain.

Table 58H. Domain Analysis of NOV58					
gnl Pfam pfam00029, connexin, Connexin. SEQ ID NO:859					
CD-Length = 218 residues, 100.0% aligned					
Score = 265 bits (678), Expect = 2e-72					
NOV58:	1	MSWMFLRDLLSGVKNKYSTGTGWIWLA	VVVFRLLVYMVAAEHVWKDEQKEFECNSRQPGC	60	
Sbjct:	1	M W FL LL GVNK+ST G IWL+V+F+FR+LV	VAAE VW DEQ +F CN++QPGC	60	
NOV58:	61	KNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAYHEGREKRRKK-----LYVSP	113		
Sbjct:	61	+NVC+D FFPIS VRLW LQLI VSTPSLL + HVAY RE++ R+K LY	120		
NOV58:	114	G-----TMDGGLWYAYLISLIVKTGF	EIGFLVLFYKLYDGFSVPYLIKCDLKPC	162	
Sbjct:	121	+ GGLW+ Y+ S+I K+ FE+GFL Y LY GF++ L+ C PC	179		
NOV58:	163	PNTVDCFISKPTTEKTIFILFLVITSCLCIVLNFIELSFL	201		
Sbjct:	180	P+TVDCF+S+PTEKTIFI+F+++ S +C++LN EL +L	218		

Connexins are a family of integral membrane proteins that oligomerise to form intercellular channels that are clustered at gap junctions. These channels are specialized sites of cell-cell contact that allow the passage of ions, intracellular metabolites and messenger molecules (with molecular weight <1-2 kD) from the cytoplasm of one cell to its apposing neighbours. They are found in almost all vertebrate cell types, and somewhat similar proteins have been cloned from plant species. Invertebrates utilise a different family of molecules, innexins, that share a similar predicted secondary structure to the vertebrate connexins, but have no sequence identity to them.

Vertebrate gap junction channels are thought to participate in diverse biological functions. For instance, in the heart they permit the rapid cell-cell transfer of action potentials, ensuring coordinated contraction of the cardiomyocytes. They are also responsible for neurotransmission at specialised 'electrical' synapses. In non-excitabile tissues, such as the liver, they may allow metabolic cooperation between cells. In the brain, glial cells are extensively-coupled by gap junctions; this allows waves of intracellular Ca²⁺ to propagate through nervous tissue, and contribute to their ability to spatially-buffer local changes in extracellular K⁺ concentration.

The connexin protein family is encoded by at least 13 genes in rodents, with many homologues cloned from other species. They show overlapping tissue expression patterns, most tissues expressing more than one connexin type. Their conductances, permeability to different molecules, phosphorylation and voltage-dependence of their gating, have been found to vary. Possible communication diversity is increased further by the fact that gap junctions may be formed by the association of different connexin isoforms from apposing cells.

However, in vitro studies have shown that not all possible combinations of connexins produce active channels.

Hydropathy analysis predicts that all cloned connexins share a common transmembrane (TM) topology. Each connexin is thought to contain 4 TM domains, with two extracellular and three cytoplasmic regions. This model has been validated for several of the family members by in vitro biochemical analysis. Both N- and C-termini are thought to face the cytoplasm, and the third TM domain has an amphipathic character, suggesting that it contributes to the lining of the formed-channel. Amino acid sequence identity between the isoforms is ~50-80%, with the TM domains being well conserved. Both extracellular loops contain characteristically conserved cysteine residues, which likely form intramolecular disulphide bonds. By contrast, the single putative intracellular loop (between TM domains 2 and 3) and the cytoplasmic C-terminus are highly variable among the family members. Six connexins are thought to associate to form a hemi-channel, or connexon. Two connexons then interact (likely via the extracellular loops of their connexins) to form the complete gap junction channel. Two sets of nomenclature have been used to identify the connexins. The first, and most commonly used, classifies the connexin molecules according to molecular weight, such as connexin43 (abbreviated to Cx43), indicating a connexin of molecular weight close to 43 kD. However, studies have revealed cases where clear functional homologues exist across species that have quite different molecular masses; therefore, an alternative nomenclature was proposed based on evolutionary considerations, which divides the family into two major subclasses, alpha and beta, each with a number of members.

Due to their ubiquity and overlapping tissue distributions, it has proved difficult to elucidate the functions of individual connexin isoforms. To circumvent this problem, particular connexin-encoding genes have been subjected to targeted-disruption in mice, and the phenotype of the resulting animals investigated. Around half the connexin isoforms have been investigated in this manner. Further insight into the functional roles of connexins has come from the discovery that a number of human diseases are caused by mutations in connexin genes. For instance, mutations in Cx32 give rise to a form of inherited peripheral neuropathy called X-linked dominant Charcot-Marie-Tooth disease. Similarly, mutations in Cx26 are responsible for both autosomal recessive and dominant forms of nonsyndromic deafness, a disorder characterised by hearing loss, with no apparent effects on other organ systems.

The disclosed NOV58 is a connexin-like protein localized to gap junctions. Gap junctions were first characterized by electron microscopy as regionally specialized structures

on plasma membranes of contacting adherent cells. These structures were shown to consist of cell-to-cell channels. Proteins, called connexins, purified from fractions of enriched gap junctions from different tissues differ. The connexins are designated by their molecular mass. Another system of nomenclature divides gap junction proteins into 2 categories, alpha and beta, according to sequence similarities at the nucleotide and amino acid levels. For example, CX43 is designated alpha-1 gap junction protein, whereas CX32 and CX26 are called beta-1 and beta-2 gap junction proteins, respectively. This nomenclature emphasizes that CX32 and CX26 are more homologous to each other than either of them is to CX43.

Willecke et al. (1990) used rat connexin gene probes in Southern blot analysis of human-mouse somatic cell hybrids to map the CX26 gene to chromosome 13. By means of somatic cell hybrids, Hsieh et al. (1991) assigned the GJB2 gene to chromosome 13 in man and chromosome 14 in the mouse. Haefliger et al. (1992) showed that the rat homologs of the CX26 and CX46 genes are tightly linked on chromosome 14. By isotopic in situ hybridization, Mignon et al. (1996) mapped GJB2 to 13q11-q12 and confirmed the assignment to mouse chromosome 14. Kelsell et al. (1997) studied a pedigree containing individuals with autosomal dominant deafness and identified a mutation in the CX26 gene: a 101T-C transition resulting in a met34-to-thr amino acid substitution. CX26 mutations resulting in premature stop codons were also found in 3 autosomal recessive nonsyndromic sensorineural deafness pedigrees, genetically linked to 13q11-q12, where the CX26 gene is localized.

Immunohistochemical staining of human cochlear cells for CX26 demonstrated high levels of expression. Kelley et al. (1998) presented evidence that the 101T-C missense mutation identified by Kelsell et al. (1997) in individuals with autosomal dominant nonsyndromic deafness is not sufficient to cause hearing loss. Carrasquillo et al. (1997) performed linkage analysis in 2 interrelated inbred kindreds in a single Israeli-Arab village containing more than 50 individuals with nonsyndromic recessive deafness. Genetic mapping demonstrated that a gene located at 13q11 (DFNB1) segregated with the deafness in these 2 kindreds. Haplotype analysis, using 8 microsatellite markers spanning 15 cM in 13q11, suggested the segregation of 2 different mutations in this extended kindred; affected individuals were homozygotes for either haplotype or compound heterozygotes. Carrasquillo et al. (1997) identified 2 distinct mutations, trp77 to arg and 35delG, in the CX26 gene, both of which were predicted to inactivate connexin 26.

The recombination of marker alleles involving polymorphisms in 13q11, at known map distances from the mutations, allowed them to estimate the age of the mutations to be 3 to 5 generations (75 to 125 years). The study demonstrated that in small populations with high

rates of consanguinity, as compared with large outbred populations, recessive mutations may have very recent origin and show allelic diversity. They pointed to the same phenomenon being observed for Hurler syndrome with 3 unique mutations and for metachromatic leukodystrophy with 5 distinct mutations, discovered among the Druze and Muslim Arab villages in Israel. In light of these findings, the authors commented that it is likely that homozygosity mapping studies in highly inbred communities may be compromised, as may be studies of mapping by linkage disequilibrium, unless the possibility of mutational diversity is taken into account.

Lench et al. (1998) studied the role of CX26 mutations in singleton (sporadic) cases of nonsyndromal sensorineural deafness. Such mutations were identified in 4 of 43 U.K. and 2 of 25 Belgian patients. Thus, about 10% of families presenting with a child sporadically affected with this disorder can be offered definitive mendelian recurrence risks. This was said to be the first genetic test available for screening such children. Kelley et al. (1998) analyzed 58 multiplex families each having at least 2 affected children diagnosed with autosomal recessive nonsyndromic deafness. Mutations in both alleles of GJB2 were observed in 20 of the 58 families. A 30delG allele occurred in 33 of the 116 chromosomes, for a frequency of 0.284. This mutation was observed in 2 of 192 control chromosomes, for an estimated gene frequency of 0.01 +/- 0.007. The homozygous frequency of the 30delG allele was then estimated at 0.0001, or 1 in 10,000. Given that the frequency of all childhood hearing impairment is 1 in 1,000 and that half of that is genetic, the specific mutation 30delG is responsible for 10% of all childhood hearing loss and for 20% of all childhood hereditary hearing loss. Six novel mutations were also observed in the affected population.

Murgia et al. (1999) studied 53 unrelated individuals with nonsyndromic sensorineural hearing impairment and carried out CX26 mutation analysis. Mutations were found in 53% of cases, in 35.3% of those in whom autosomal recessive inheritance was thought likely and in 60% of the presumed sporadic cases. Three novel mutations were found. The hearing deficit varied from mild to profound even within the same family. Among patients with profound hearing loss, 35.5% were found to have a mutation; among those severely impaired, 20%; and among those moderately impaired, 33.3%.

Rabionet et al. (2000) analyzed the GJB2 gene in 576 families/unrelated patients with recessive or sporadic deafness from Italy and Spain, 193 of them being referred as autosomal recessive and the other 383 as apparently sporadic. Of the 1,152 unrelated GJB2 chromosomes, 37% had GJB2 mutations. A total of 23 different mutations were detected. Mutation 35delG was the most common, accounting for 82% of all GJB2 deafness alleles. It

represented 88% of the alleles in Italian patients and only 55% in Spanish cases. Sobe et al. (2000) sequenced the entire coding region of the GJB2 gene in 75 hearing-impaired children and adults in Israel. Was both prelingual and postlingual, with hearing loss ranging from moderate to profound. Almost 39% of all persons tested harbored GJB2 mutations, most of which were 35delG and 167delT. A novel mutation, involving both a deletion and an insertion, 51del12insA, was identified in a family originating from Uzbekistan. All GJB2 mutations were associated with prelingual hearing loss, although severity ranged from moderate to profound, with variability even among hearing-impaired sibs. No significant difference in hearing levels was found between individuals with 35delG and 167delT mutations.

Antoniadi et al. (2000) screened 26 unrelated Greek patients with prelingual sensorineural deafness in whom syndromic forms and environmental causes of deafness had been excluded. They detected the 35delG mutation in 28 chromosomes (53.8%); another 3 sequence variations accounted for 7.6% of the alleles. Wilcox et al. (2000) performed mutation analysis of the GJB2 gene and audiology on 106 families presenting with at least 1 child with congenital hearing loss. In 74 families (80 children), the etiology was consistent with nonsyndromic recessive hearing loss. Six different GJB2 mutations, including 1 novel mutation, were identified. They found that GJB2 mutations caused a range of phenotypes from mild to profound hearing impairment and that loss of hearing in the high-frequency range (4,000 to 8,000 Hz) is a characteristic feature in children with molecularly diagnosed CX26 hearing impairment. They also demonstrated that high frequency hearing loss was found in a group of similar size of deaf children in whom a mutation could be found in only one of the GJB2 alleles. In their study, the M34T mutation was associated with hearing loss only when present in compound heterozygous state, suggesting autosomal recessive inheritance.

Heathcote et al. (2000) reported a missense mutation in affected members of a family with autosomal dominant deafness and palmoplantar keratoderma. Rabionet et al. (2000) reviewed the molecular genetics of hearing impairment due to mutations in gap junction genes encoding beta-connexins. Among these genes, mutations in GJB2 account for about 50% of all congenital cases of hearing impairment. Three mutations in GJB2 are particularly common in specific populations: 35delG in Caucasoids, 167delT in Ashkenazi Jews, and 235delC in Orientals. Carrier frequencies in these populations vary between 1 and 30 and 1 in 75. Over 50 mutations have been identified in the GJB2 gene, of which some missense changes have a dominant-negative action in hearing impairment, with partial to full penetrance. Functional studies for some missense mutations in connexins 26, 30, and 32 indicate abnormal gap junction conductivity. Expression patterns in mouse and rat cochlea indicate that connexin 26

and connexin 30 are expressed in the supporting cells of the cochlea, suggesting a potential role in endolymph potassium recycling.

In the Japanese population, Kudo et al. (2000) sequenced the GJB2 gene in 39 patients with prelingual deafness, 39 patients with postlingual progressive sensorineural hearing loss, and 63 individuals with normal hearing. GJB2 mutations were found in 5 of the 39 patients (12%) with prelingual deafness. The most common mutation was 235delC observed in 7 of 10 mutant alleles. There were no cases with the 30delG allele. No GJB2 mutation was found in patients in the postlingual hearing loss group. Nance et al. (2000) noted that recessive mutations at the connexin-26 gene locus account for nearly half of all cases of genetic deafness in many populations. They suggested that this high frequency is only seen in populations with a long tradition of intermarriage among deaf people. Available data are consistent with the hypothesis that such marriages might well have contributed to the high frequency of connexin-26 deafness in the U.S., and could represent a novel mechanism for maintaining specific genotypes at unexpectedly high frequencies.

The NOV58 disclosed in this invention is predicted to be expressed in at least the following tissues: brain, lung, ovary, and colon. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV58 is provided in Example 2.

The nucleic acids and proteins of NOV58 are useful in potential therapeutic applications implicated in various gap junction-related pathological disorders described further herein. The NOV58 nucleic acid encoding the connexin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a Connexin-like protein includes the nucleic acid whose sequence is provided in Table 58A or 58C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 58A or 58C while still encoding a protein that maintains its connexin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence shown in Table 58A or 58C, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids

whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 29% of the bases may be so changed. The novel protein of the invention includes the connexin-like protein whose sequence is provided in Table 58B or 58D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 58B or 58D while still encoding a protein that maintains its connexin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 44% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV59

A eukaryotic translation initiation factor 5 (EIF5), disclosed herein as NOV59, interacts with the 40S initiation complex to promote hydrolysis of bound GTP with concomitant joining of the 60S ribosomal subunit to the 40S initiation complex. The resulting functional 80S ribosomal initiation complex is then active in peptidyl transfer and chain elongations. The disclosed NOV59 (alternatively referred to herein as CG56633-01) includes the 1328 nucleotide sequence (SEQ ID NO:213) shown in Table 59A. A NOV59 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 34-36 and ends with a TGA codon at nucleotides 1273-1275. The disclosed NOV59 maps to human chromosome 3.

Table 59A. NOV59 Nucleotide Sequence (SEQ ID NO:213)

```
CTTTCCTCATCACCTTAAATTCGGGTGTCCTTTATGAGTAATCAAAAGCAGCAAAAGCCA
ACGCTATCAGGCCAGTATTTAAACCAGAAAAAGAGATGAAAAAGAGAGGTTTGACCCCT
ACTCAGTTTCAGGACTACGTTATTCAAGGCTTAAGTGAAGTGGTACTGATTTGGAAGCA
GTAGCTAAGTTCCTTGATGCTTCTGGAACAAAAGTATTACCGTCGATGTGCAGAAACA
CTCTTTGACATTCTGGTGGGTGGTGAATGCTGGCCCCAGGTGGTACACTGGCAGATGAC
ATCATGCGTACAGATGTCTGCGTGTTCAGGCCAAGAAGACCTAGAGACCATGCAAGCA
TTTGCTCAGGTTTTTAACAAGTTAATCAGGCACTACAAATACCTTGAGAAATGTTGTGAA
GATGAAGTAAAAAGGCTGCTGGTGTTCGGAAGGGTTTTTCAGACTCGGAGAGGAAAAAA
CTGGCTATGTTGACTGGTGTCTCTGGCTAATGCATCCATTCTTAATAGCCTTTATAAT
GAGAATTTGGTTAAAGAAGGGGTTCAACAGCTTTTGCTGGAAGCTATTTAAATCATGT
ATAAATGAAAAAGATATCAATGCAGTAAGTCAAGGAAAGTCAGCATGGATAACAGCCTG
ATGGAACCTTTTCTGCAATAAGCAAAGCGTTCAACACTTCACGAAGTATTTTACTGAG
GCAGGCCTGAAAGAGCTTTCAGAATATGTTTCGGAATCAGCAAACCATCAGAGCTTGTAA
```

GAGCTGCAGAAAGAACTTCAAGAACAGATGTCCCGTGGGGATCCATTTAAGGTTATAATT
TTATATGTCAAGGAGGAGATGAAAAAACAACATCCCAGAACAGTTGTCATCGAAATA
GTCTGGTCAAATGTAATGAGCGCTGTGGAATGGAACAAAAGAGAGGAGATTGTAGCAGAG
CAAGCCATCAAACACTTGAAGCAACACAGCCCTCTACTTGCTGCCTTTACTACTCAAAGT
CAGTCTGAGCTGACCCTGTTACTGAAGATTGAGGAGTATTGCTATGACAACATTCATTTC
ATGAAAGCCTTACGGAAAATAGTGCTGCTTTTATAAAGCTGTAGTCCTGAGCAAAGAG
ACCATTTTGAAGTGGTATAAAGGTACACATGTTGCAAAGGGGAAGAGTGTTTTCCTTGAG
CAAATGAAAAAGTTTGAGAGTGGCTCAAAATGCTGAAGAAGATCTGAATCTGAAGCT
GAAGAAGGTGACTGAATTTGAACTACACCTCAGTAAAGCAAACAGGAGTTGTAGATA
AAATGTCC

The NOV59 polypeptide (SEQ ID NO:214) encoded by SEQ ID NO:213 is 413 amino acids in length and is presented using the one-letter amino acid code in Table 59B. The Psort profile for NOV59 predicts that this sequence has no signal peptide and is likely to be localized to the nucleus with a certainty of 0.7600. In alternative embodiments, a NOV59 polypeptide is located to lysosomes with a certainty of 0.10000.

Table 59B. NOV59 Polypeptide Sequence (SEQ ID NO:214)

MSNQKQQKPTLSGPVFKTRKRDEKERFDPTQFQDYVIQGLTETGTDLEAVAKFLDASGTK
LDYRRCATLFDIILVGGMLAPGGTLADDIMRTDVCVFAAQEDLETMQAFAQVFNKLI RH
YKYLEKCCEDVVKRLLVFGKGFSDSERKKLAML TGVL LANASILNSLYNENLVKEGVSTA
FAGKLFKSCINEKDINAVTARKVSMDSLMELFPANKQSVQHFTKYFTEAGLKELSEYVR
NQQTIRACKELQKELQEOMSRGDPFKVIIILYVKEEMKKNNIPEPVVIEIVWSNVMSAVEW
NKREEIVAEQAIKHLKQHSPLLAFTTQSQSELTLLLKIQEYCYDNIHFMKALRKIVLFL
YKAVVLSKETILKWKYKTHVAKGKSVFLEQMKKFGEWLKNAAEESESEAAEEGD

A BLAST analysis of NOV59 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV59 had high homology to other proteins as shown in Table 59C.

Table 59C. BLASTX results from PatP database for NOV59

Smallest		High	Sum
		Score	Probability P (N)
Sequences producing High-scoring Segment Pairs:			
patp:AAB43883	Human cancer associated protein sequence	1834	5.6e-189
patp:AAW93950	Human regulatory molecule HRM-6 protein	1403	2.6e-143
patp:AAB92726	Human protein sequence	1403	2.6e-143
patp:AAM38764	Human polypeptide	1403	2.6e-143
patp:AAM40550	Human polypeptide	1403	2.6e-143

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 778 of 824 bases (94%) identical to a gb:GENBANK-ID:HUMRSC419|acc:D13630.1 mRNA from *Homo sapiens* (Human mRNA for KIAA0005 gene. The full amino acid sequence of the protein of the invention was found to have 372 of

419 amino acid residues (88%) identical to, and 385 of 419 amino acid residues (91%) similar to, the 419 amino acid residue ptrn:SPTREMBL-ACC:Q15394 protein from *Homo sapiens* (Human) (KIAA0005 PROTEIN). NOV59 also has homology to the other proteins shown in the BLASTP data in Table 59D.

5

Table 59D. NOV59 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7661850 ref NP_055485.1 (NM_014670)	basic leucine-zipper protein BZAP45; KIAA0005 gene product [<i>Homo sapiens</i>]	419	372/419 (88)	385/419 (91)	0.0
gi 7661744 ref NP_054757.1 (NM_014038)	HSPC028 protein [<i>Homo sapiens</i>]	419	264/406 (65)	355/406 (82)	e-143
gi 15341786 gb AAH13060.1 AAH13060 (BC013060)	HSPC028 protein [<i>Mus musculus</i>]	419	264/406 (65)	334/406 (82)	e-143
gi 4426565 gb AAD20436.1 (AF031483)	unknown [<i>Rattus norvegicus</i>]	419	264/406 (65)	334/406 (82)	e-143
gi 11640562 gb AAG39278.1 AF110323_1 (AF110323)	MSTP017 [<i>Homo sapiens</i>]	419	263/406 (64)	334/406 (81)	e-143

This BLASTP data is displayed graphically in the ClustalW in Table 59E. A multiple sequence alignment is given, with the NOV59 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 59D.

10

Table 59E. ClustalW Alignment of NOV59					
NOV59	(SEQ ID NO:214)				
gi 7661850	(SEQ ID NO:597)				
gi 7661744	(SEQ ID NO:598)				
gi 15341786	(SEQ ID NO:599)				
gi 4426565	(SEQ ID NO:600)				
gi 11640562	(SEQ ID NO:601)				
<div> <div>102030405060</div> <div> <div>NOV591</div> <div>gi 7661850 1</div> <div>gi 7661744 1</div> <div>gi 15341786 1</div> <div>gi 4426565 1</div> <div>gi 11640562 1</div> </div> <div> <div>MSNQKQOKPTLSGPVFKTRKRDEKERFDPTQFQDYVIOGLTETGTDLEAVAKFLDASGTR</div> <div>MNNQKQOKPTLSGORFKTRKRDEKERFDPTQFQDCILIOGLTETGTDLEAVAKFLDASGAK</div> <div>MN--KHQKPVLTGQRFKTRKRDEKEKFEPFVFRDTLVQGLNEAGDDLEAVAKFLDSTGSR</div> <div>MN--KHQKPVLTGQRFKTRKRDEKEKFEPFVFRDTLVQGLNEAGDDLEAVAKFLDSTGSR</div> <div>MN--KHQKPVLTGQRFKTRKRDEKEKFEPFVFRDTLVQGLNEAGDDLEAVAKFLDSTGSR</div> <div>MN--KHQKPVLTGQRFKTRKRDEKEKFEPFVFRDTLVQGLNEAGDDLEAVAKFLDSTGSR</div> </div> </div>					
<div> <div>708090100110120</div> <div> <div>NOV5961</div> <div>gi 7661850 61</div> <div>gi 7661744 59</div> </div> <div> <div>LDYRRCAETLFDILVAGCMLAPGGTLADD---IMRTDVCVFANQEDLETMQAFAQVFNKL</div> <div>LDYRRYAETLFDILVAGCMLAPGGTLADD---MMRTDVCVFANQEDLETMQAFAQVFNKL</div> <div>LDYRRYADTLFDILVAGSMLAPGGTRIDDGDKTKMTHCVFSANEDHETIRNYAQVFNKL</div> </div> </div>					

gi	15341786	59	LDYRRYADTLFDILVAGSMLAPGGTRIDDGDKTKMTNHCVFSANEDHETIRNYAQVFNKL	118
gi	4426565	59	LDYRRYADTLFDILVAGSMLAPGGTRIDDGDKTKMTNHCVFSANEDHETIRNYAQVFNKL	118
gi	11640562	59	LDYRRYADTLFDILVAGSMLAPGGTRIDDGDKTKMTNHCVFSANEDHETIRNYAQVFNKL	118
<div> <div>130140150160170180</div> <div> <div>NOV59</div> <div>gi 7661850</div> <div>gi 7661744</div> <div>gi 15341786</div> <div>gi 4426565</div> <div>gi 11640562</div> </div> <div> <div>118</div> <div>118</div> <div>119</div> <div>119</div> <div>119</div> </div> <div> <div>IRRYKYLEKCCCEDEVKRLLVFGKGFSDSERKKLAMLICVLLAN----</div> <div>IRRYKYLEKCFEDEVKRLLLFLKGFSESERKKLAMLICVLLANGTLNASILNSLYNENLV</div> <div>IRRYKYLEKAFEDEMKKLLFLKAFSETEQTKLAMLSGILLGNGTLPATILTSLFTDSL</div> <div>IRRYKYLEKAFEDEMKKLLFLKAFSETEQTKLAMLSGILLGNGTLPATILTSLFTDSL</div> <div>IRRYKYLEKAFEDEMKKLLFLKAFSETEQTKLAMLSGILLGNGTLPATILTSLFTDSL</div> </div> <div> <div>173</div> <div>177</div> <div>178</div> <div>178</div> <div>178</div> </div> </div>				
<div> <div>190200210220230240</div> <div> <div>NOV59</div> <div>gi 7661850</div> <div>gi 7661744</div> <div>gi 15341786</div> <div>gi 4426565</div> <div>gi 11640562</div> </div> <div> <div>174</div> <div>178</div> <div>179</div> <div>179</div> <div>179</div> </div> <div> <div>KEGVSTAFACKLFGKSCINEKDINAVTA--RKVSMDSNLMELFPANKQSVQHEHTKYFTEAG</div> <div>KEGVSAFAVAVKLFKSWINEKDINAVASLRKVSMDNRLMELFPANKQSVQHEHTKYFTEAG</div> <div>KEGIAASFAVAVKLFKAWMAEKDANSVTSSLRKANLDKRLELFPVNRQSVDFHAKYFTDAG</div> <div>KEGIAASFAVAVKLFKAWMAEKDANSVTSSLRKANLDKRLELFPVNRQSVDFHAKYFTDAG</div> <div>KEGIAASFAVAVKLFKAWMAEKDANSVTSSLRKANLDKRLELFPVNRQSVDFHAKYFTDAG</div> </div> <div> <div>231</div> <div>237</div> <div>238</div> <div>238</div> <div>238</div> </div> </div>				
<div> <div>250260270280290300</div> <div> <div>NOV59</div> <div>gi 7661850</div> <div>gi 7661744</div> <div>gi 15341786</div> <div>gi 4426565</div> <div>gi 11640562</div> </div> <div> <div>232</div> <div>238</div> <div>239</div> <div>239</div> <div>239</div> </div> <div> <div>LKELSEYVRNQOTTIRACKELQKELQEQMSRGDPFKVITILVYVKEEMKKNNIPFPVVIETVW</div> <div>LKELSEYVRNQOTTICARKELQKELQEQMSRGDPFKDTILVYVKEEMKKNNIPFPVVIETVW</div> <div>LKELSDFLRVQOSLGRKELQKELQERLSQECPIKEVVLYVKEEMKRNDLPETAVIGLLW</div> <div>LKELSDFLRVQOSLGRKELQKELQERLSQECPIKEVVLYVKEEMKRNDLPETAVIGLLW</div> <div>LKELSDFLRVQOSLGRKELQKELQERLSQECPIKEVVLYVKEEMKRNDLPETAVIGLLW</div> </div> <div> <div>291</div> <div>297</div> <div>298</div> <div>298</div> <div>298</div> </div> </div>				
<div> <div>310320330340350360</div> <div> <div>NOV59</div> <div>gi 7661850</div> <div>gi 7661744</div> <div>gi 15341786</div> <div>gi 4426565</div> <div>gi 11640562</div> </div> <div> <div>292</div> <div>298</div> <div>299</div> <div>299</div> <div>299</div> </div> <div> <div>SNVMSAVEWNKKEELVAEQALKHLKQYAPLLAVFSSQGSSELVLLQKVQVEYCYDNIHFMK</div> <div>SSVMSAVEWNKKEELVAEQALKHLKQYAPLLAVFSSQGSSELVLLQKVQVEYCYDNIHFMK</div> <div>TCIMNAVEWNKKEELVAEQALKHLKQYAPLLAVFSSQGSSELVLLQKVQVEYCYDNIHFMK</div> <div>TCIMNAVEWNKKEELVAEQALKHLKQYAPLLAVFSSQGSSELVLLQKVQVEYCYDNIHFMK</div> <div>TCIMNAVEWNKKEELVAEQALKHLKQYAPLLAVFSSQGSSELVLLQKVQVEYCYDNIHFMK</div> </div> <div> <div>351</div> <div>357</div> <div>358</div> <div>358</div> <div>358</div> </div> </div>				
<div> <div>370380390400410420</div> <div> <div>NOV59</div> <div>gi 7661850</div> <div>gi 7661744</div> <div>gi 15341786</div> <div>gi 4426565</div> <div>gi 11640562</div> </div> <div> <div>352</div> <div>358</div> <div>359</div> <div>359</div> <div>359</div> </div> <div> <div>ALRKIVVLFYKAVVLSKETILKWKYKGTTHVAKGKSVFLEQMKKFEWLNAAEESESEAE</div> <div>AFQKIVVLFYKAEVLSSEETILKWKYKGTTHVAKGKSVFLEQMKKFEWLNAAEESESEAE</div> <div>AFQKIVVLFYKADVLSEETILKWKYKGTTHVAKGKSVFLEQMKKFEWLNAAEESESEAE</div> <div>AFQKIVVLFYKADVLSEETILKWKYKGTTHVAKGKSVFLEQMKKFEWLNAAEESESEAE</div> <div>AFQKIVVLFYKADVLSEETILKWKYKGTTHVAKGKSVFLEQMKKFEWLNAAEESESEAE</div> </div> <div> <div>411</div> <div>417</div> <div>418</div> <div>418</div> <div>418</div> </div> </div>				
NOV59	412	GD	413	
gi 7661850	418	GD	419	
gi 7661744	419	N	419	
gi 15341786	419	S	419	
gi 4426565	419	S	419	
gi 11640562	419	N	419	

Table 59F lists the domain description from DOMAIN analysis results against NOV59. This indicates that the NOV59 sequence has properties similar to those of other proteins known to contain this domain.

Table 59F. Domain Analysis of NOV59

gnl Load LOAD W2, W2, conserved protein-protein interaction domain in translation factors like eIF2B SEQ ID NO: 860					
CD-Length = 116 residues, 96.6% aligned					
Score = 83.6 bits (205), Expect = 2e-17					
NOV59:	290	VWSNVMSAVEWNKREEIVAEQA	IKHLKQHSPLLAFTTQSQSELTLL	LKIQEYCYDNIHF	349
Sbjct:	1	VALVILSVASIELADNEPKEAAVKVFKKWGPLLAKY	LKDEDSQLELLYAL	EEFC EELEEL	60
NOV59:	350	MKALRKIVVLFYKAVVLSKETILKWY-KGTHVAKGKSVFLEQM	KKFGEWLKN		400
Sbjct:	61	LKLLAKILKYLYDEDVLEEEAILKWYEKKSKAEEGKKKVLKSAKPFVTWLQ	E		112

The NOV59 disclosed in this invention is predicted to be expressed in at least the following tissues: brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV59 is provided in Example 2.

The nucleic acids and proteins of NOV59 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, myasthenia gravis, neuroprotection, endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders and other diseases, disorders and conditions of the like. The NOV59 nucleic acid encoding the translation initiation factor 5-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a translation initiation factor 5-like protein includes the nucleic acid whose sequence is provided in Table 59A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 59A while still encoding a protein that maintains its translation initiation factor 5-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures

include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 6% of the residues may be so changed.

The novel protein of the invention includes the translation initiation factor 5-like protein whose sequence is provided in Table 59B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 59b while still encoding a protein that maintains its translation initiation factor 5-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 12% of the bases may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV60

NOV60 includes two Lynx1-like proteins, designated herein as NOV60a and NOV60b, which differ by three amino acids and the relative length of their untranslated regions (UTR's).

NOV60a

The disclosed NOV60a (alternatively referred to herein as CG56894-01) includes the 715 nucleotide sequence (SEQ ID NO:215) shown in Table 60A. A NOV60a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 348-350 and ends with a stop codon at nucleotides 696-698.

Table 60A. NOV60a Nucleotide Sequence (SEQ ID NO:215)

```
AGCTTTGTTCTTGAGTGGGTCTGCCTCGGGGGCTTTAGAGGAGACCCAGAGGGTGGCGATGCGGCACGG
GTGCTGCGGGACACACAGACACGCCTACGATTAGACTCAGGCAGGCACCTACCGGCAGAGCGGCGCGGGT
GACTCCAGGCGCGCGGTACCTCACGGTGGTGAAGGTCACAGGGTTGCAGCACTCCAGTAGACCAGGA
GCTCCGGGAGGCAGGGCCGGCCCCACGTCTCTGCGCACCACCCTGAGTTGGATCCTCTGTGCGCCACCC
CTGAGTTGGATCCAGGGCTAGCTGCTGTGACCTCCCCACTCCCACGCTGCCCTCCTGCCTGCAGCCATG
ACGCCCTTGCTCACCTGATCCTGGTGGTCCTCATGGGCTTACCTCTGGCCCAGGCCTCGGACTGCCACG
TGTGTGCCTACAACGGAGACAACCTGCTTCAACCCCATGCGCTGCCCGGCTATGGTTGCCTACTGCATGAC
CACGCGCACCTACTACACCCCAACAGGATGAAGGTCAGTAAGTCCCTGCGTGCCCCGCTGCTTCGAGACT
GTGTATGATGGCTACTCCAAGCACGCGTCCACCACCTCCTGCTGCCAGTACGACCTCTGCAACGGCACCG
GCCTTGCCACCCCGGCCACCCCGGCCCTGGCCCCATCCTCCTGGCCACCCTCTGGGGTCTCCTCTAAAG
CCCCGAGGCAGACA
```

The NOV60a polypeptide (SEQ ID NO:216) encoded by SEQ ID NO:215 is 116 amino acids in length and is presented using the one-letter amino acid code in Table 60B. The Psort profile for NOV60a predicts that this sequence has a signal peptide and is likely to be exported from the cell with a certainty of 0.8200. In alternative embodiments, a NOV60a polypeptide is located to lysosomes with a certainty of 0.1000, or to the endoplasmic reticulum (membrane) with a certainty of 0.1000. The Signal P predicts a likely cleavage site for a NOV60a peptide is between positions 34 and 35, *i.e.*, at the dash in the sequence AQA-SD.

Table 60B. NOV60a Polypeptide Sequence (SEQ ID NO:216)

MTPLLTLLILVVLMLGLPLAQASDCHVCAYNNGDNCFNPMRCPAMVAYCMTTRTYTPTRMKV
SKSCVPRCFETVYDGYSKHASTTSCCQYDLNCGTGLATPATPALAPILLATLWGLL

NOV60b

The disclosed NOV60b (alternatively referred to herein as CG56894-02) includes the 876 nucleotide sequence (SEQ ID NO:) shown in Table 60C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 348-350 and ends with a stop codon at nucleotides 696-698.

Table 60C. NOV60b Nucleotide Sequence (SEQ ID NO:217)

AGCTTTGTTCTTGAGTGGGTCTGCCTCGGGGGCTTTAGAGGAGACCCAGAGGGTGGCGA
TGCGGCACGGGTGCTGCGGGACACAGACACGCTTACGATTAGACTCAGGCAGGCACCT
ACCGGCGAGCGGCGCGGGTGACTCCAGGCGCGGCGGTACCTCAGGTGGTGAAGGTCA
CAGGGTTGCAGCACTCCAGTAGACCAGGAGCTCCGGGAGGCAGGGCCGCCCCACGTCC
TCTGCGCACCACTTGAAGTGGATCCTCTGTGCGCCACCCCTGAGTTGGATCCAGGGCTA
GCTGCTGTTGACCTCCCACTCCACGCTGCCCTCCTGCCTGCAGCATGACGCCCTGC
TCACCCTGATCCTGGTGGTCTCATGGGCTTACCTCTGGCCCAGGCCTCGGACTGCCACG
TGTGTGCCTACAACGGAGACAAGTCTTCAACCCCATGCGCTGCCCGGCTATGGTTGCCT
ACTGCATGACCACGCGACCTACTACACCCCAACAGGATGAAGGTCAAGTACCTCGG
TGCCCCGCTGCTTCGAGACTGTGTATGATGGCTACTCCAAGCACGCGTCCACCACCTCCT
GCTGCCAGTACGACCTCTGCAACGGCACCGGCTTGCCACCCCGGCCACCTGGCCCTGG
CCCCATCCTCCTGGCCACCTCTGGGGTCTCCTCTAAAGCCCCCGAGGCAGACCCACTC
AAGAACAAGCTCTCGAGACACACTGCTACACCTCGCACCCAGCTCACCTGCCTCACC
CTCCACACTCCTGCGACCTCCTCAGCATGCCAGGGTCAGGACTGTGGGCAAGAAGAC
ACCGGACCTCCCCCAACCAACACACGACCTCACTTC

The NOV60b polypeptide (SEQ ID NO:218) encoded by SEQ ID NO:217 is 116 amino acids in length and is presented using the one-letter amino acid code in Table 60D. The Psort profile for NOV60b predicts that this sequence is a Type Ia membrane protein, has a signal peptide, and is likely to be localized at the plasma membrane with a certainty of 0.9190. In alternative embodiments, a NOV60b polypeptide is located to lysosomes with a certainty of 0.2000, or to the endoplasmic reticulum (membrane) with a certainty of 0.1000. The Signal P

predicts a likely cleavage site for a NOV60b peptide is between positions 20 and 21, *i.e.*, at the dash in the sequence AQA-SD.

Table 60D. NOV60b Polypeptide Sequence (SEQ ID NO:218)

MTPLLTLILVVLMLPLAQASDCHVCAYNGDNCFNPMRCPAMVAYCMTTRTYTPTRMKV
SKSCVPRCFETVYDGYSKHASTTSCCQYDLCNGTGLATPATLALAPILLATLWGLL

5

A BLAST analysis of NOV60 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV60 had high homology to other proteins as shown in Table 60E.

Table 60E. BLASTX results from PatP database for NOV60

		High Score	Smallest Sum Probability P (N)
Sequences producing High-scoring Segment Pairs:			
patp:AAY02738	Human secreted protein encoded by gene 89	630	2.2e-61
patp:AAM39828	Human polypeptide	630	2.2e-61
patp:AAM41614	Human polypeptide	630	2.2e-61
patp:AAB61131	Human NOV3 protein	594	1.4e-57
patp:AAY79325	Mouse receptor ligand Lynx1	521	7.7e-50

10

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 338 of 424 bases (79%) identical to a gb:GENBANK-ID:AF141377|acc:AF141377.1 mRNA from *Mus musculus* (Ly-6/neurotoxin homolog (Lynx1) mRNA). The full amino acid sequence of the protein of the invention was found to have 92 of 116 amino acid residues (79%) identical to, and 96 of 116 amino acid residues (82%) similar to, the 116 amino acid residue ptnr:SPTREMBL-ACC:Q9WVC2 protein from *Mus musculus* (Mouse) (LY-6/NEUROTOXIN HOMOLOG). NOV60 also has homology to the other proteins shown in the BLASTP data in Table 60F.

15

Table 60F. NOV60 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7106349 ref NP_035968.1 (NM_011838)	Ly6/neurotoxin 1 [<i>Mus musculus</i>]	116	92/116 (79%)	96/116 (82%)	5e-39

gi 12698684 gb AAK01642.1 AF321824_1 (AF321824)	Ly-6 neurotoxin-like protein Lynx1 [Homo sapiens]	80	79/80 (98%)	79/80 (98%)	5e-27
gi 1519481 gb AAB07524.1 (U66837)	E48 antigen [Homo sapiens]	79	28/72 (38%)	34/72 (46%)	0.035
gi 10720241 sp P57096 PSCA_MOUSE E48 antigen [Homo sapiens]	E48 antigen [Homo sapiens]	123	30/104 (28%)	40/104 (37%)	0.038
gi 12845967 dbj BAB26976.1 (AK010485)	PAR/Ly-6 domain containing protein [Mus musculus]	154	36/108 (33%)	47/108 (43%)	0.068

This BLASTP data is displayed graphically in the ClustalW in Table 60G. A multiple sequence alignment is given, with the NOV60a and b protein being shown on lines 1 and 2 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 60F.

Table 60G. ClustalW Alignment of NOV60									
NOV60a	(SEQ ID NO:216)								
NOV60b	(SEQ ID NO:218)								
gi 7106349	(SEQ ID NO:602)								
gi 12698684	(SEQ ID NO:603)								
gi 1519481	(SEQ ID NO:604)								
gi 10720241	(SEQ ID NO:605)								
gi 12845967	(SEQ ID NO:606)								
		10	20	30	40	50	60		
NOV60a	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
NOV60b	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
gi 7106349	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
gi 12698684	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
gi 1519481	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
gi 10720241	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
gi 12845967	1	MTPLLTLLVVL	MGLPLAQASDCHVCAYNG--DNCFNEMRCPAMVAYCMTTRTYTPT	57					
		70	80	90	100	110	120		
NOV60a	58	MKVS	KSCVPRCFET-VYDGYSKHAS	102					
NOV60b	58	MKVS	KSCVPRCFET-VYDGYSKHAS	102					
gi 7106349	58	MKVS	KSCVPRCFET-VYDGYSKHAS	102					
gi 12698684	58	MKVS	KSCVPRCFET-VYDGYSKHAS	102					
gi 1519481	40	NLVE	KDCAESCTPSYTLQGLVSSGTS	79					
gi 10720241	60	TVIS	KGCSQCEDD--SENYYLGKK	103					
gi 12845967	59	FYVS	KQCTRRCPPTVVSPPSPNPPSEPKEFLIEKPMFPLYKCCQWDSCNGEGPPTDQLL	118					
		130	140	150					

NOV60a	103	ALAP-----	ILLATLWGILL-----	116
NOV60b	103	ALAP-----	ILLATLWGILL-----	116
gi 7106349	103	ALVP-----	ALLATFWSILL-----	116
gi 12698684	80	-----	-----	80
gi 1519481	79	-----	-----	79
gi 10720241	104	TTLG-----	LLTVLCSILLWGSSRL--	123
gi 12845967	119	KEQPGKASGRRHRYIEL	LLTGFMVITANGLSALCLL	154

Elapid snake venom neurotoxins exert their effects through high-affinity interactions with specific neurotransmitter receptors. The lynx1-like gene disclosed herein as NOV60, is highly expressed in the brain and contains the cysteine-rich motif characteristic of this class of neurotoxins. Primary sequence and gene structure analyses reveal an evolutionary relationship between lynx1 and the Ly-6/neurotoxin gene family. Lynx1 is expressed in large projection neurons in the hippocampus, cortex, and cerebellum. In cerebellar neurons, lynx1 protein is localized to a specific subdomain including the soma and proximal dendrites. Lynx1 binding to brain sections correlates with the distribution of nAChRs, and application of lynx1 to *Xenopus* oocytes expressing nAChRs results in an increase in acetylcholine-evoked macroscopic currents. These results identify NOV60 as a protein modulator for nAChRs in vitro, with important implications in the regulation of cholinergic function in vivo.

The NOV60 disclosed in this invention is predicted to be expressed in at least the following tissues: brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV60 is provided in Example 2.

The nucleic acids and proteins of NOV60 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, and neurodegeneration, as well as other diseases, disorders and conditions. The NOV60 nucleic acid encoding the lynx1-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The nucleic acid of the invention encoding a lynx1-like protein includes the nucleic acid whose sequence is provided in Table 60A or 60C, or a fragment thereof.

The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 60A or 60C while still encoding a protein that maintains its Lynx1-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 60A or 60C including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 21% of the bases may be so changed.

The novel protein of the invention includes the lynx1-like protein whose sequence is provided in Table 60B and 60D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 60B and 60D while still encoding a protein that maintains its lynx1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 21% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV61

NOV61 is a homolog of the adlcan gene and belongs to the superfamily of cell adhesion molecules. The disclosed NOV61 (alternatively referred to herein as CG56453-01) includes the 5925 nucleotide sequence (SEQ ID NO:) shown in Table 61A. A NOV61 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 16-18 and ends with a stop codon at nucleotides 5653-5655. The disclosed NOV 70 maps to human chromosome Y.

**Table 61A. NOV61 Nucleotide Sequence (SEQ ID
NO:219)**

AGGCACCCCGACAAGATGCCCAAGCGCGCGCACTGGGGGGCCCTCTCTGTGGTGCTGATC

CTGCTTTGGGGTTCATCCGCGAGTGGCGCTGGCCTGCCCTCATCCTTGTGCCTGCTACGTC
 CCCAGCGAGGTCCACTGCACGTTCCGATCCCTGGCTTCTGTGCCCGCTGGCATTGCTAAA
 CATGTGGAAGAATCAATTTGGGGTTTGGAAATCTGAAGTGTAAGGACAAAGCTTAT
 GAAGGCGGTCAAGTTGTGTGCAATGTGCTTCAGTCCAAAGAAGTTGTACAAACATGAGATT
 CACAAGCTGAAGGACCTGACTTGTCTGAAGCCTTCCATAGAGTCTCCTCTGAGACAGAAC
 AGGAGCAGGAGTATTGAGGAGGAGCAAAACAAGAAGAGAATGGTGACAGCCAGCTCATC
 CTGGAGAAAAATCCAACCTCCCCAGTGGAGCATCTCTTTGAATATGACTGATGAGCACGGG
 AACCTGGTGAACCTGGTGTGTGACATCAAGAAACCAATGGATGTGTACAAAATTCACCTG
 AACCAAACAGATCCTCCAGATATTGACATAAATGCAATGGTTGCCTTGGACTTTGAGTAT
 AGTATGACCCAGGAAAACTATGAAAATCTATGGAATTTGATAGCATACTACAGTGAAGTT
 CCCATGAAGCTACACAGAGAGCTCATGCTCAGCAAACACCCAGAGTCAGCTACCAGTAC
 AGGCAAGATCCCGATGAAGAAGCTCTTACTACACAGGTGTGAGAGCCAGATTCTTGCA
 GAACCAAGATGGATCATGCAGCCATCCATAGATATCCAGCTGAACCGACCTCAGAGTACG
 GCCAAGAAGGTGCTACTTTCTACTACAACAGTATTCTCAAACAATAGCCACCAAGAT
 ACAAGGCGAGTCCGGGGCAGAAGCTGGGTAATGATTGAGCCTAGTAGAGCTGTGCAAAAA
 GATCAGACTGTCTTGGAAAGGGGTCGATGCCAGTTGAGCTGCAATGTGAAAGCTTCTGAG
 AGTCCATCTATCTTCTGGGTGCTTCCAGATGGCTCCATCCTGAAAGTGCCTGTGGATGAC
 CCAGACAGCAAGTTCTCCATTCTCAGCAGTGGCTGGCTGAGGATCAAGTCCATGGAGCCA
 TCTGACTCGGGCTTGTACCAAGTGCATTGCTCAAGTGAGGGATGAAATGGACCGCATGGTA
 TATAGGGTACTTGTGCACTCTCCCTCCACTCAGCCAGCCGAGAAAGACACAGTGACAAT
 GGCAAGAACCAGGGGAGCCAGTGTGTTGCCTTGCAATGCTTTAGCTATACCCGAAGCC
 CACCTTAGCTGGATTCTTCCAAACAGAAGGATAATTAATGATTGGCTAACACATCACAT
 GTATACATGCTGCCAAATGGAACCTTTCCATCCCAAAGGTCCAAGTCAGTGACAGTGGT
 TACCACAGATGTGTGGCTGTCAACCAGCATGGGGCAGACCATATCACGGTGGGAATCACA
 GTGACCAAGAAAGGTTCTGGCTCGCCATCCAAAGAGGAGAGATGGCCAGGTCCAAAGGCT
 CTTTCCAGAATGAGAGAAGACATCGTGGAGGATGAAGGGGTCTCAGGCACGGGAGATGAA
 GAGAACACTTCAAGGAGACTTCTACATCCAAAGCACCAGAGGGCTTCTCAAACCAAG
 GATGATGCCATCAATGGAGATAAGAAAGCCAAAGAAAGGAGAAGAAAGCTGAATCTGG
 AAGCATTTCAGAAAAAGAACAGAGACAGTGTGTCAGAAAGATCTCAGAGTGTGTAATCA
 AGACGAAGGATAAACGTGGCAACAAACAGATTAATCCGGAGCACTGGGCTGATATTTTA
 GCCAAAGTCTTTGGGAAAAATCTCCCTACAGGCACAGAAGTATCCCAATTTATAAAACC
 ACAAGTTCTCCATTCTTGAGCCTAGTAGTCACACCACCTTTGCCTGCTGTTTCTCCCCC
 TTGGCATCTCCAATACAGACAGCAACAAGTGTGAAGAATCCTCAGCAGATGTACCTCTA
 CTCAGCGAAGGAAAGCACATTTTGTAGTACCATTTCTCAGCCAGCATGGGACTAGAACAC
 CACAACAATGGAGTTATTCTTGTGAACCTGAAGTAACAAGCACACCTCTGGAAGAAGTT
 GTTGATGAGTATTCCAAGAAGACTGAGGAGATGACTTCCACTGAAGGCGAGCTGAAGGGG
 ACTGCAGCCTCTACACTTATATCTGAGCCTTATGAACAATCTCCTACTCTACACACCTTA
 GACACAGTCTATGAAGAGCCACCCATGAAGAGACGGAACAGAGGTTGGTCTGCAGCA
 GATGTGGATCTCACCAGATCCACATCCAGTGAAGTATGAGCTTCCATTGGTTGTTGTG
 TCTTGGCTGAGTCTAAGCCTGTGCAATACTTTGACCCAGATTGGAGACTAATTCACAA
 CCACATGAGGATAACATAAAGAATACAGTTTGCACACCTTACTCCAACCGCCATCATC
 TGGTTTAATGACTCTAGTACATCACTGTCTATTGAGGATTCTACTGTAGGGGAACAAGGT
 TGCCAGGCAAAATCACATCTACAAGGACCGACAGAGAACATCCAGCTTGTGAAAAGTAGT
 TTTAGCACTCAAGACACCTTATTGATTAAAAAAGGTATGAAAGAGATGTCTCAGACACTA
 CAGGGAGGAAATATGTAGAGGGAGACCTACACACTCCAGAAGTTCTGAGAATGAGGGC
 CAAGAGAGCAAAATCCATCACTTACCTGACTCCACACTGGGTATAACGAGCAGTACGCTCT
 CCAGTTAAGAAGCCTGCGGAAACACAGTTGTCAACCTGCTACACAAAGACACACA
 GAAACAACCTCAAGGCAAAAAGTGGCTTCATCATCCACCATGAGCACTCACCTTCTCGA
 AGGAGACCCAATGGGAGAAAAATTACACCTCACAAATTCACCACCGGCACAAGCAAAACC
 CCACCCACAACCTTTGTCTCATTAGAGACTTTTCTACTCAACCAACTCAAGCAACTGAC
 ATTAAGATTTCAAATCAAATGGAGAGTTCTCTGGTTCTTACATCTTGGGAGATTAAACACA
 GTTAATACCCCCAACAGCTGGAAATGGAGAAGAATGTAGAGCTCATATCAAAGGGAAC
 CCACGGAGAAAACACGGGAAGAGGCCAAACAAACATCGATATACCCCTTCTACAGTGAAGT
 TCAAGAGCATCTGCATCCAAGCCAGCCCTTCTCCAGAAAATAAACATAGAAACATTGTT
 ACTCCAGTTCAGAACTACACTTTTGCCTAGAAATGTTTCTCTGAAAACCTGAGGGCGTT
 TATGATTCTTAGATTACACGACAACCACGAAAAATACATTCTCACCATAAAGTC
 CAAGACACACTTCCAGTCATGTATAAACCCACATCAGATGGAAAAAGAAATTCAGGATGAT
 GTTGCCACAAATGTTGACAAACATAAAGTGACATTTTAGTCCCTGGTGAGTCAATTACA
 AATGTACACAAACTTCTCGCTCCTTGGTCTCCACTATGGGAGAATTTAAGGAAGATCC
 TCTCCTGTGGGCTTTCCAGGAATTTCAACCTGGAATCCCTCAAGGAAAGCTCAGCCTGGG
 AGGCTACAGACAGACATACATGTTACCCTTCTGGGGAAACCCCTACAGACCTTCCCTT
 TTTAACGAGCTTGAGGATGTGGATTTTACTTCTGAGTTTGTGCTCTGTGACAGTCTCC
 ACACCATTTCAACAGGAAGAAGCTGGTTTTTCCACAATTCTCTCAAGCATAAAGTGGAG
 ATGGCTTCAAGTCAGGTAGAACTACCACCTTGGTCAAGATCATCATGAAACCACTGTG
 GCTATTCTCACTCTGAACTAGACCACAGAATCACATCCTTACTGCTGCCTGGATGAAG
 GAGCAGCATCTTTGTGCTCCATGATTTCTCTGTCTTTGGGACAAACCCACCACT
 AAGCCAGAACTTCTCAGTCCAAGAACATCTCAAATATGTAAAGATTCCAAGGAAATGTT
 TTCTTGAATTACATGGGGAATCCAGAAACAGAAGCAACCCAGTGAAAAATGAAGGAACA
 CAGCGTATGTCAGGGCCAAATGAATTATCAACACCATCTTCTGACCACGATGCATTTAAC

TTGTCTACAAAGCTAGAATTGGAAAAGCAAGTATTTGATAGTAGGAGTCTAACACGTGGC
 CCAGATAGCCACCACCAGGATGGAAGAGTTTCATGCTTCTCATCAACTAACAGAAATCCCT
 GCCAAACCCATCCTACCAACAGGAACAGTGAGGCTGCCTGAAATGTCCACACAAAGCACT
 TCCAGATACTTTGTAACCTTCCAGCCACCTCATCACGGGACCAACAAACCAGAAATAACT
 ACATATCCTTCTAGGGCTTTGCCAGAGAGCAAACAGTTTACAACCTCCAAGAGTAGCAAGT
 ACAACTCCTTCTCTATCACACATGTCCAAACCAGCATTCTAGTAAGTTTGGCTGACCTA
 AGAACTGACCAATCCAATGGCTCCTACAAAGTGTGGAAATAGCAACATCCCTGAGGCA
 AGAAACTCAGTTGGAAAGCCTCTCAGTCCAAGAATTTATCATTATCCAATGGAAGACTC
 CCTTTCTTTACCAACAGGACTCTTTCTTTTTCACAGTTGGGAGTCACCCGGAGACCCAG
 ATACCTCTTCTCCTGTCCCAGTAATGAGAGAGAGAAAAGTTAATCCAGGTTCCCTACAAT
 AGGATATATTTCCATAGCACCTTCCATCTGGACTTTGGCCTTCCAGCACCTCCACTGTTG
 CACACTCCATGGACCATGGTATCACCCCAACTAAGTTACAGAATATCCCTATGGTCTCA
 TCCACCCAGAGTTCTGTCTCCTTTATAACATCTTCTGTCCAGTCTCAGGAAGCATCCAC
 CAAAGCGGCTCAAAGTTCTTTGCAGGAGGACCGCTGCATCCAAATTTCTGGCCTTTGGG
 GAAAGCCCCAAATCCTACCAAGTCCCCACAGACTGTGTCTGTCTGCTGAAACGGAC
 GCTGTGTTCCCGTGTGAGGCAATAGGAAAACCAAAGCCTTTCGTTACTTGGACAAAAGTT
 TCCACAGGAGTTCTTATGACTCCGAATACCAGGATACAACGGTTTGAGGTTCTCAAGAAC
 GGTACCTTAGTGATAAGGAAGTTTCAAGTGCAAGATCGAGGCCAGTATATGTGCACCGCC
 AGCAACCTGTACGGCCTGGACAGGATGGTGGTCTTTCTCTGGGTACCGTGCAGCAACCT
 CAAATCCTAGCCTCCCCTACCAGGACGTACCGTCTACCTGGGAGACACATTACAATG
 GAGTGTCTGGCGAAAGGACCCCGCCCCCAAAATTTCTGGATCTTCCGTGACAGGAG
 GTGTGGCAAACCTCTGTCTCCGTGGAGGGCCGGATCACCTGCACCAAAACCGGACCTT
 TCCATCAAGGAGGCGTCTTCTCAGACAGAGGCGTCTATAAGTGCCTGGCCAGCAACGCA
 ACCCGGGCGACAGCGTGTCCATCCGCTACACGTGGCGGCACTGCCCCCATTATCCAC
 CAGGAGAAGCTGTAGAACATCTCGCTGCCCGGGGCTCAGCATTACATTCACTGCACT
 GCCAAAGCTGCGCCCCGTCAGCGTGTCTGGGTGCTCGGGGATGGTACCCAAATCCGC
 CCTCGCATTTCTTCCACCGGAACCTGTTTGTGTTTCCCAACGGGACGCTCTACATCTGC
 AACCTCGCGCCCAAGGACAGCGGGCGCTATGAGTGCCTGGCCGCCAACCTGATCGGCTCC
 GCGCGCAGTACGGTGCAGCTGAACGTGCAGCGCAGCAGCGAAC

The NOV61 polypeptide (SEQ ID NO:220) encoded by SEQ ID NO:219 is 1879 amino acids in length and is presented using the one-letter amino acid code in Table 61B. The Psort profile for NOV61 predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.4371. In alternative embodiments, a NOV61 polypeptide is located to lysosomes with a certainty of 0.1900, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the nucleus with a certainty of 0.1800. The Signal P predicts a likely cleavage site for a NOV61 peptide is between positions 26 and 27, *i.e.*, at the dash in the sequence ALA-CP.

Table 61B. NOV61 Polypeptide Sequence (SEQ ID NO:220)

MPKRAHWGALSVVLILLWGHPRVALACPHPCACYVPSEVHCTFRSLASVPAGIAKHVERI
 NLGFGILKCKDKAYEGGQLCAMCFSPKKLYKHEIHKLDLTCLKPSIESPLRQNRSRSI
 EEEQKQEENGDSQLILEKIQLPQWSISLNMDEHGNLVNLVCDIKKPMVDVYKIHNLQTD
 PDIDINAMVALDFEYPMTOENYENLWKLIAYYSEVPMKHLHRELMLSKHPRVSYQYRQD
 EEALYYTGVRQILAEPEWIMQPSIDIQLNRPQSTAKKVLLSYNQYSQTIATKDRQAR
 GRSWVMIEPSRAVQKDQTVLEGGRCQLSCNVKASESPSIFWVLPDGSILKVPVDDPDSKF
 SILSSGWLRIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDVTITGKNP
 EPVMLPCNALAIPEAHLISWILPNRRIINDLANTSHVYMLPNGTSLIPKVQVSDSGYHRCV
 AVNQHGADHITVGITVTKKSGSGPSKRGRWPGPKALSRMREDIVEDEGVSGTGEENTSR
 RLLHPKHQEAFLKTKDDAINGDKAKKGRRLKLWKHSEKEPETSVAEDLRFESRRRIN
 VANKQINPEHWADILAKVFGKNLPTGTEVSPIIKTSSPFLSLVVTPLPAVSPPLASPI
 QTATSAEESADVPLLSEGHILSTISSASMGLEHHNNGVILVEPEVTSTPLEEVVDEYS
 KKTEEMTSTEGDLKGTAASTLISEPYEQSPTLHTLDTVYEEPTHEETETEGWSAADVGSS
 PDPTSSEYELPLVVVSLAESKPVQYFDPDLETNSQPHEDNKEYSFAHLTPTAIIWFNDS
 STLSLFEDSTVGEQGVPGKSHLQGPTENIQLVKSSFSTQDTLLIKKGMKEMSQTLLQGGNM
 LEGDPTHSRSENEGQESKITLPDSTLGTSTSTSPVKKPAETTVVTLHLKDTTETTPR
 QKVASSSTMSTHPSRRRPNGRKLHPKFKHHRHKQTPPTTFAPLETFTSTQPTQATDIKISN

QMESSLVPTSWEINTVNTPKQLEMEKNVELISKGTPRRKHGKRPNKHRYTPSTVSSRASA
SKPSPSPENKHRNIVTPSSETLLPRNVSLKTEGVYDSLDTTTRKIHSSHHKVQDTLP
VMYKPTSDGKEIQDDVATNVDKHKSDILVPGESITNVTQTSRSLVSTMGEFKEESSVPGF
PGIPTWNP SRKAQPGRLQTDIHVTTSGETPTDPPLVNELEDVDFITSEFLSSVTVPFHQ
EEAGFSTILSSIKVEMASSQVETTTLGQDHHETTVAI LHSETRPQNHILTAAMWKEPASL
SPPMILLSLGQTTTTPKELLSPRTSQICKDSKENVFLNYMGNPETEATPVKNEGTQRMMSG
PNELSTPSSDHDADFNLSTKLELEKQVFDSSRLTRGPDSSHQDGRVHASHQLTRIPAKPIL
PTGTVRLPEMSTQSTSRFYVTFQPPHHGTNKPEITTYPSRALPESKQFTTTRVASTTPLL
SHMSKPSISSKFADLRTDQSNQSYKVFNGNSNIPEARNSVGKPLSPRIYHYSNGRLPFFTN
RTLFSFSQLGVTRRPQIPSSPVPVMRERKVNPGSYNRIYSHSTFHLDGFLPAPPLLHTPWT
MVSPTNLQNI PMVSSSTQSSVSFITSSVQSSGSIHQSGSKFFAGGPPASKFWPLGKEKQI
LTKSPQTVSVTAETDAVFPCEAIGKPKPFVTWTKVSTGVLMTPNTRIQRFEVLKNGTLVI
RKQVQDRGQYMC TASNLYGLDRMVVFLWVTVQQPQILASHYQDVTVYLGDTITMECLAK
GTPAPQISWIFRDRRVQTLSSVEGRITLHQNRTLISKEASFSDRGVYKCVASNATRADS
VSIRLHVAALPPIIHQEKL

A BLAST analysis of NOV61 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV61 had high homology to other proteins as shown in Table 61C.

5

Table 61C. BLASTX results from PatP database for NOV61		
Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)
patp:AAM03157 Peptide #1839 encoded by probe	2631	2.0e-273
patp:AAM15395 Peptide #1829 encoded by probe	2631	2.0e-273
patp:AAM27883 Peptide #1920 encoded by probe	2631	2.0e-273
patp:AAM55191 Human brain expressed single exon probe	2631	2.0e-273
patp:AAM67586 Human bone marrow expressed probe	2631	2.0e-273

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 4045 of 4330 bases (93%) identical to a gb:GENBANK-ID:AF245505|acc:AF245505.1 mRNA from *Homo sapiens* (adlican mRNA). The full amino acid sequence of the protein of the invention was found to have 1598 of 1818 amino acid residues (87%) identical to, and 1661 of 1818 amino acid residues (91%) similar to, the 2828 amino acid residue ptnr:SPTREMBL-ACC:Q9NR99 protein from *Homo sapiens* (Human) (ADLICAN). NOV61 also has homology to the other proteins shown in the BLASTP data in Table 61D.

15

Table 61D. NOV61 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14766612 ref XP_035465.1 (XM_035465)	hypothetical protein XP_035465 [<i>Homo sapiens</i>]	2828	1590/1818 (87)	1654/1818 (90)	0.0

gi 9280405 gb AAF86402.1 A F245505_1 (AF245505)	Adlican [Homo sapiens]	2828	1591/1818 (87)	1654/1818 (90)	0.0
---	---------------------------	------	-------------------	-------------------	-----

This BLASTP data is displayed graphically in the ClustalW in Table 61E. A multiple sequence alignment is given, with the NOV61 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 61D.

Table 61E. ClustalW Alignment of NOV61					
NOV61	(SEQ ID NO:220)				
gi 14766612	(SEQ ID NO:607)				
gi 9280405	(SEQ ID NO:608)				
		10	20	30	40
NOV61	1	MPKRAHWGALS	VVILLWGHPRVALACPHCACYVPSEVHCTFRSLASVPAGIAKHVERI	60	
gi 14766612	1	MPKRAHWGALS	VVILLWGHPRVALACPHCACYVPSEVHCTFRSLASVPAGIAKHVERI	60	
gi 9280405	1	MPKRAHWGALS	VVILLWGHPRVALACPHCACYVPSEVHCTFRSLASVPAGIAKHVERI	60	
		70	80	90	100
NOV61	61	NLGFNSIQALSETSFAGLTKLELLMIHGNEIPSIDGALRDLSSLQVFKFSYNKLRVITG	63		
gi 14766612	61	NLGFNSIQALSETSFAGLTKLELLMIHGNEIPSIDGALRDLSSLQVFKFSYNKLRVITG	120		
gi 9280405	61	NLGFNSIQALSETSFAGLTKLELLMIHGNEIPSIDGALRDLSSLQVFKFSYNKLRVITG	120		
		130	140	150	160
NOV61	63	QTLOGLSNLMRLHIDHNKIEFIHPQAFNGLTSLRLLHLEGNLLHQLHPSTFTFTFLDYF	64		
gi 14766612	121	QTLOGLSNLMRLHIDHNKIEFIHPQAFNGLTSLRLLHLEGNLLHQLHPSTFTFTFLDYF	180		
gi 9280405	121	QTLOGLSNLMRLHIDHNKIEFIHPQAFNGLTSLRLLHLEGNLLHQLHPSTFTFTFLDYF	180		
		190	200	210	220
NOV61	64	RLSTIRHLYLAENMVRTLPAASLRNMPLENLYLQGNPWTCDCEMRWFLEWDAKSRGILK	68		
gi 14766612	181	RLSTIRHLYLAENMVRTLPAASLRNMPLENLYLQGNPWTCDCEMRWFLEWDAKSRGILK	240		
gi 9280405	181	RLSTIRHLYLAENMVRTLPAASLRNMPLENLYLQGNPWTCDCEMRWFLEWDAKSRGILK	240		
		250	260	270	280
NOV61	69	CKKDKAYEGGQLCAMCFSPKKLYKHEIHKLKDMTCLKPSIESPLRQNRSRSEIEEQE	128		
gi 14766612	241	CKKDKAYEGGQLCAMCFSPKKLYKHEIHKLKDMTCLKPSIESPLRQNRSRSEIEEQE	300		
gi 9280405	241	CKKDKAYEGGQLCAMCFSPKKLYKHEIHKLKDMTCLKPSIESPLRQNRSRSEIEEQE	300		
		310	320	330	340
NOV61	129	DGGSQILILEKFQLPQWSISLNMTEHGNMNVNLVCDIKKPMVDYKIHNLQTDPPDIDINAT	188		
gi 14766612	301	DGGSQILILEKFQLPQWSISLNMTEHGNMNVNLVCDIKKPMVDYKIHNLQTDPPDIDINAT	360		
gi 9280405	301	DGGSQILILEKFQLPQWSISLNMTEHGNMNVNLVCDIKKPMVDYKIHNLQTDPPDIDINAT	360		
		370	380	390	400
NOV61	189	VALDFECPMTRENYEKLWKLIAYYSEVPVKLHRELMLSKDPRVSQYQRQDADEEALYYTG	248		
gi 14766612	361	VALDFECPMTRENYEKLWKLIAYYSEVPVKLHRELMLSKDPRVSQYQRQDADEEALYYTG	420		
gi 9280405	361	VALDFECPMTRENYEKLWKLIAYYSEVPVKLHRELMLSKDPRVSQYQRQDADEEALYYTG	420		
		430	440	450	460
NOV61	249	VRAQILAEPWVMQPSIDIQLNRRQSTAKKVLLSYNQYSQTIETKTDTRQARGRSWVMIE	308		
gi 14766612	421	VRAQILAEPWVMQPSIDIQLNRRQSTAKKVLLSYNQYSQTIETKTDTRQARGRSWVMIE	480		

gi 9280405	421	VRAQILAEPEWVMQPSIDIQLNRRQSTAKKVLLSYTQYSQTISTKDTRQARGRSWVMIE	480
		490 500 510 520 530 540	
NOV61	309	PSRAVQKDQTVLEGGRCQLSCNVKASESPSIFWVLPDGSILKVFMDPPDSKFSILSSGWL	368
gi 14766612	481	PSGAVORDQTVLEGGPCQLSCNVKASESPSIFWVLPDGSILKAPMDPPDSKFSILSSGWL	540
gi 9280405	481	PSGAVORDQTVLEGGPCQLSCNVKASESPSIFWVLPDGSILKAPMDPPDSKFSILSSGWL	540
		550 560 570 580 590 600	
NOV61	369	RIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDVTVTIGKNPGEVMLPCN	428
gi 14766612	541	RIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDVTVTIGKNPGESVTLPCN	600
gi 9280405	541	RIKSMEPSDSGLYQCIAQVRDEMDRMVYRVLVQSPSTQPAEKDVTVTIGKNPGESVTLPCN	600
		610 620 630 640 650 660	
NOV61	429	ALAIPEAHLNWILPNRRIINDLANTSHVYMLPNGTSLIPKVQVSDSGYHRCVAVNQHIGAD	488
gi 14766612	601	ALAIPEAHLNWILPNRRIINDLANTSHVYMLPNGTSLIPKVQVSDSGYYRCVAVNQQGAD	660
gi 9280405	601	ALAIPEAHLNWILPNRRIINDLANTSHVYMLPNGTSLIPKVQVSDSGYYRCVAVNQQGAD	660
		670 680 690 700 710 720	
NOV61	489	HITVGITVTKKGSGSPSKRGRWPGPKALSRMRREDIVEDEGVSGTGDEENTSRRLHHPKHQ	548
gi 14766612	661	HFTVGITVTKKGSGLPSKRGRRPGAKALSRVREDIVEDEGGSGMGDEENTSRRLHHPKDQ	720
gi 9280405	661	HFTVGITVTKKGSGLPSKRGRRPGAKALSRVREDIVEDEGGSGMGDEENTSRRLHHPKDQ	720
		730 740 750 760 770 780	
NOV61	549	EAFLLTKDDAINGDKKAKKGRRLKLVKHSEKEPETSVAEIDLRFESRRRINVANKQINP	608
gi 14766612	721	EVFLTKDDAINGDKKAKKGRRLKLVKHSEKEPETNVAEGRRVFESRRRINMANKQINP	780
gi 9280405	721	EVFLTKDDAINGDKKAKKGRRLKLVKHSEKEPETNVAEGRRVFESRRRINMANKQINP	780
		790 800 810 820 830 840	
NOV61	609	EHWADILAKVFGKNLPTGTEVSPFIKTTSSPFLSLVVTPLPAVSPPLASPIQTATSAAE	668
gi 14766612	781	ERWADILAKVRGKNLPKGTEVPPLIKTTSPPSLSLEVTPPPFAVSPPSASPVQTVTSAAE	840
gi 9280405	781	ERWADILAKVRGKNLPKGTEVPPLIKTTSPPSLSLEVTPPPFAVSPPSASPVQTVTSAAE	840
		850 860 870 880 890 900	
NOV61	669	SSADVPLLSECKHILSTISSASMGLEHNHNGVILVEPEVTSTPLEEVVDLSEKTEEITS	728
gi 14766612	841	SSADVPLLGEHHVLTGTEVSPFIKTTSSPFLSLVVTPLPAVSPPLASPIQTATSAAE	900
gi 9280405	841	SSADVPLLGEHHVLTGTEVSPFIKTTSSPFLSLVVTPLPAVSPPLASPIQTATSAAE	900
		910 920 930 940 950 960	
NOV61	729	TEGDLKGTAASTLISEPYEQSPTLHTLDTVYEEPTHEETETEGWSAADVGSSPDPTSSEY	788
gi 14766612	901	TEGDLKGTAAPTLLISEPYEPSPTLHTLDTVYEEKPTHEETATGWSAADVGSSPEPTSSEY	960
gi 9280405	901	TEGDLKGTAAPTLLISEPYEPSPTLHTLDTVYEEKPTHEETATGWSAADVGSSPEPTSSEY	960
		970 980 990 1000 1010 1020	
NOV61	789	ELPLVVVSLAESKPVQYFDPDLETNSQPHEDNIKEYSFAHLTPTAIWFNDSSSTLSLFED	848
gi 14766612	961	EPPLDAVSLAESEPMQYFDPDLETKSQPDDEKMKEDTFAHLTPPTIWFNDSSSTLSLFED	1020
gi 9280405	961	EPPLDAVSLAESEPMQYFDPDLETKSQPDDEKMKEDTFAHLTPPTIWFNDSSSTLSLFED	1020
		1030 1040 1050 1060 1070 1080	
NOV61	849	STVGEQGVPGKSHLQGLTDNIHLVKSSLSSTQDTLLIKKGMKEMSQTLOGGNMLEGDPHTS	908
gi 14766612	1021	STIGEPGVPGQSHLQGLTDNIHLVKSSLSSTQDTLLIKKGMKEMSQTLOGGNMLEGDPHTS	1080
gi 9280405	1021	STIGEPGVPGQSHLQGLTDNIHLVKSSLSSTQDTLLIKKGMKEMSQTLOGGNMLEGDPHTS	1080
		1090 1100 1110 1120 1130 1140	
NOV61	909	RSESEGOESKSITLPDSTLGISSSPVKKPAETTVTLLHKDTTTTETTTPRQKVASSS	967
gi 14766612	1081	RSESEGOESKSITLPDSTLGISSSPVKKPAETTVGTLDDKDTTTATTTPRQKVAPSS	1140
gi 9280405	1081	RSESEGOESKSITLPDSTLGISSSPVKKPAETTVGTLDDKDTTTATTTPRQKVAPSS	1140
		1150 1160 1170 1180 1190 1200	

NOV61	968	TMSTHPSRRRPNGRKLHPKFFHRRHKQTPPTTFAPLETFSTOPTQATDIKISNOVESSL	1026
gi 14766612	1141	TMSTHPSRRRPNGRRRLRPNKFRHRHKQTPPTTFAPSETFSTOPTQADIKISSQVESSL	1200
gi 9280405	1141	TMSTHPSRRRPNGRRRLRPNKFRHRHKQTPPTTFAPSETFSTOPTQADIKISSQVESSL	1200
		1210 1220 1230 1240 1250 1260	
NOV61	1027	VPTSWELNTVNTPKQLEMEKNVLELISKGTPRRKHGKRPKNHRYTPSTVSSRASASKPSPS	1086
gi 14766612	1201	VPTAWVDNTVNTPKQLEMEKNAEPTSKGTPRRKHGKRPKNHRYTPSTVSSRASGSKPSPS	1260
gi 9280405	1201	VPTAWVDNTVNTPKQLEMEKNAEPTSKGTPRRKHGKRPKNHRYTPSTVSSRASGSKPSPS	1260
		1270 1280 1290 1300 1310 1320	
NOV61	1087	PENKHRNIVTPSSETILLPRNVSLKTEGVYDSLDYITTTTRKIHSSHHKVDITLPVMYKPT	1146
gi 14766612	1261	PENKHRNIVTPSSETILLPRTVSLKTEGVYDSLDYMTTTRKIYSSYPKVQETLPVTKPT	1320
gi 9280405	1261	PENKHRNIVTPSSETILLPRTVSLKTEGVYDSLDYMTTTRKIYSSYPKVQETLPVTKPT	1320
		1330 1340 1350 1360 1370 1380	
NOV61	1147	SDGKEIKDDVATNVDKHKSDILVPGESITNVTOISRLVSTMGEFKEESSPVGFPGIPTW	1206
gi 14766612	1321	SDGKEIKDDVATNVDKHKSDILVPGESITNAIPTSRSLVSTMGEFKEESSPVGFPGTPTW	1380
gi 9280405	1321	SDGKEIKDDVATNVDKHKSDILVPGESITNAIPTSRSLVSTMGEFKEESSPVGFPGTPTW	1380
		1390 1400 1410 1420 1430 1440	
NOV61	1207	NPSRKAQPGRLQTDIHVTTSGETPTDPLVNELEDVDFTSEFLSSVTVPFHFQEEAGFS	1266
gi 14766612	1381	NPSRTAQPGRLOTGIPVTTSGENLTDPELLKELEDVDFTSEFLSSLTVPFHFQEEAGSS	1440
gi 9280405	1381	NPSRTAQPGRLOTDIPVTTSGENLTDPELLKELEDVDFTSEFLSSLTVPFHFQEEAGSS	1440
		1450 1460 1470 1480 1490 1500	
NOV61	1267	TTLSSIKVEVASSQVETTTLCODHHEHTTVAILHSETRPQNHILTAAMKEPASLSPMIL	1326
gi 14766612	1441	TTLSSIKVEVASSQAETTTLDQDHLETTVAILLSETRPQNHPTAARMKEPASSSPSTIL	1500
gi 9280405	1441	TTLSSIKVEVASSQAETTTLDQDHLETTVAILLSETRPQNHPTAARMKEPASSSPSTIL	1500
		1510 1520 1530 1540 1550 1560	
NOV61	1327	LSLGQTTTTPKPELSPRISCLCKDSKENVFLNYMGNPETEATPVKNEGTORMSGPNELST	1386
gi 14766612	1501	MSLGQTTTTPKALPSPRISQASRDSKENVFLNYVGNPETEATPVNNEGTOHMSGPNELST	1560
gi 9280405	1501	MSLGQTTTTPKALPSPRISQASRDSKENVFLNYVGNPETEATPVNNEGTOHMSGPNELST	1560
		1570 1580 1590 1600 1610 1620	
NOV61	1387	PSSDQDAFNLSTKLELEKQVFGSRSLPRGPDSDGRVHASHQLTRVPAKPILPTATVR	1446
gi 14766612	1561	PSSDQDAFNLSTKLELEKQVFGSRSLPRGPDSDGRVHASHQLTRVPAKPILPTATVR	1620
gi 9280405	1561	PSSDQDAFNLSTKLELEKQVFGSRSLPRGPDSDGRVHASHQLTRVPAKPILPTATVR	1620
		1630 1640 1650 1660 1670 1680	
NOV61	1447	LPEMSTQSTSRFYVTFQPPHFGTNKPEITTYPSRALPESKQFTTPRVASTT-PLLSHMSK	1505
gi 14766612	1621	LPEMSTQASRYFVTSQSPRHWTNKPEITTYPSGALPENKQFTTPRLSSTTIPLPLHMSK	1680
gi 9280405	1621	LPEMSTQASRYFVTSQSPRHWTNKPEITTYPSGALPENKQFTTPRLSSTTIPLPLHMSK	1680
		1690 1700 1710 1720 1730 1740	
NOV61	1506	PSISSKFADLRDQNGSYKVFGNSNIPEARNSVGKPLSPRIVHYNSGRLPFFTNKTLSF	1565
gi 14766612	1681	PSIPSKFTDRRTDQFNGYSKVFGNNNIPEARNPVGKPPSPRIPHYNSGRLPFFTNKTLSF	1740
gi 9280405	1681	PSIPSKFTDRRTDQFNGYSKVFGNNNIPEARNPVGKPPSPRIPHYNSGRLPFFTNKTLSF	1740
		1750 1760 1770 1780 1790 1800	
NOV61	1566	SQLGVTRRPQIPSSPVPMRERKVNPGSYNRIYSHSTFHLDGFLPAPPLLHTPQTGSPS	1625
gi 14766612	1741	PQLGVTRRPQIPTSPAPVMRERKVI PGSYNRIHSHSTFHLDGPPAPPLLHTPQTGSPS	1800
gi 9280405	1741	PQLGVTRRPQIPTSPAPVMRERKVI PGSYNRIHSHSTFHLDGPPAPPLLHTPQTGSPS	1800
		1810 1820 1830 1840 1850 1860	
NOV61	1626	TNLQNI PMVSSSTQSSVSFITSSVQSSGSIHQSGSKFFAGGPPASKFWPLGKPKQILT KSP	1685
gi 14766612	1801	TNLQNI PMVSSSTQSSVSFITSSVQSSGSIHQSGSKFFAGGPPASKFWSLGKPKQILT KSP	1860

gi 9280405	1801	TNLQNI PMVSSSTQSSISFITSSVQSSGSFHQSSSKFFAGGPPASKFWSLGEKQPILTKSP	1860
		1870 1880 1890 1900 1910 1920	
NOV61	1686	QTVSVTAETDAVFPCEATGKPKPFVTWTKVSTGVLMTPNTRIQRFEVLKNGTLVIRKQOV	1745
gi 14766612	1861	QTVSVTAETD TVFPCEATGKPKPFVTWTKVSTGALMTPNTRIQRFEVLKNGTLVIRKQOV	1920
gi 9280405	1861	QTVSVTAETD TVFPCEATGKPKPFVTWTKVSTGALMTPNTRIQRFEVLKNGTLVIRKQOV	1920
		1930 1940 1950 1960 1970 1980	
NOV61	1746	QDRGQYMCTASNLHGLDRMVVLLSVTVQQPQILASHYQDVTVYLGDTIAMECLAKGTPAP	1805
gi 14766612	1921	QDRGQYMCTASNLHGLDRMVVLLSVTVQQPQILASHYQDVTVYLGDTIAMECLAKGTPAP	1980
gi 9280405	1921	QDRGQYMCTASNLHGLDRMVVLLSVTVQQPQILASHYQDVTVYLGDTIAMECLAKGTPAP	1980
		1990 2000 2010 2020 2030 2040	
NOV61	1806	QISWIFDRRRVWQTVSPVEGRITLHENRTLSIKEASFSDRGVYKCVASNAAGADSLAIRL	1865
gi 14766612	1981	QISWIFDRRRVWQTVSPVEGRITLHENRTLSIKEASFSDRGVYKCVASNAAGADSLAIRL	2040
gi 9280405	1981	QISWIFDRRRVWQTVSPVESRITLHENRTLSIKEASFSDRGVYKCVASNAAGADSLAIRL	2040
		2050 2060 2070 2080 2090 2100	
NOV61	1866	HVAALPPVIHQEKL-----	1879
gi 14766612	2041	HVAALPPVIHQEKL ENISLPPGLSIHIHCTAKAAPLSVRWVLGDGTQIRPSQFLHGNLF	2100
gi 9280405	2041	HVAALPPVIHQEKL ENISLPPGLSIHIHCTAKAAPLSVRWVLGDGTQIRPSQFLHGNLF	2100
		2110 2120 2130 2140 2150 2160	
NOV61	1879	-----	1879
gi 14766612	2101	VFPNGTLYIRNLAPKDSGRYECVAANLVGSARRTVQLNVQRAAANARITGTSPPRTDVRY	2160
gi 9280405	2101	VFPNGTLYIRNLAPKDSGRYECVAANLVGSARRTVQLNVQRAAANARITGTSPPRTDVRY	2160
		2170 2180 2190 2200 2210 2220	
NOV61	1879	-----	1879
gi 14766612	2161	GGTLKLDCSASGDPWPRILWRLP SKRMIDALFSFDSRIKVFANGTLVVKSVTDKDGADYL	2220
gi 9280405	2161	GGTLKLDCSASGDPWPRILWRLP SKRMIDALFSFDSRIKVFANGTLVVKSVTDKDGADYL	2220
		2230 2240 2250 2260 2270 2280	
NOV61	1879	-----	1879
gi 14766612	2221	CVARNKVGDDYVVLKVDVVMKPAKIEHKEENDHKVFYGGDLKVD CVATGLPNPEISWSLP	2280
gi 9280405	2221	CVARNKVGDDYVVLKVDVVMKPAKIEHKEENDHKVFYGGDLKVD CVATGLPNPEISWSLP	2280
		2290 2300 2310 2320 2330 2340	
NOV61	1879	-----	1879
gi 14766612	2281	DGSLVNSFMQSDSGGRTKRYVFNNGTLYFNEVGMREEGDTYCF AENQVGKDEMVRVK	2340
gi 9280405	2281	DGSLVNSFMQSDSGGRTKRYVFNNGTLYFNEVGMREEGDTYCF AENQVGKDEMVRVK	2340
		2350 2360 2370 2380 2390 2400	
NOV61	1879	-----	1879
gi 14766612	2341	VVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTWLSPTNKVIPT SSEKYQIYQD	2400
gi 9280405	2341	VVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTWLSPTNKVIPT SSEKYQIYQD	2400
		2410 2420 2430 2440 2450 2460	
NOV61	1879	-----	1879
gi 14766612	2401	GTL LIQKAQRSDSGNYTCLVRNSAGEDRKT VVIHVNVPKPKINGNPNPITTVREIAAGGS	2460
gi 9280405	2401	GTL LIQKAQRSDSGNYTCLVRNSAGEDRKT VVIHVNVPKPKINGNPNPITTVREIAAGGS	2460
		2470 2480 2490 2500 2510 2520	
NOV61	1879	-----	1879
gi 14766612	2461	RKLIDCKAEGIPTPRVLWAFPEGVVLPA PYYGNRITVHNGSLDIRSLRKSDSVQLVCMA	2520
gi 9280405	2461	RKLIDCKAEGIPTPRVLWAFPEGVVLPA PYYGNRITVHNGSLDIRSLRKSDSVQLVCMA	2520
		2530 2540 2550 2560 2570 2580	

is a protein that has been described to be elevated in patients with osteoarthritis. Sequence analysis indicates that it is likely to be a secreted protein. A rat gene named mechanical stress-induced protein has been patented and shows significant similarity to adlcan. This protein is elevated in osteoblasts subjected to mechanical stress and has been suggested to be effective in the prognosis, diagnosis or treatment of osteoarthritis. Since this family of proteins seems to have involvement in osteoarthritis, it follows that the protein of invention may share that characteristic.

The disclosed NOV61 of invention has two significant attributes - it is truncated relative to its homolog, adlcan, and secondly, it maps to the Y chromosome. The first attribute is significant in that it is possible that the truncated adlcan-like protein may play a dominant-negative role in the function of adlcan. It is therefore possible that it may constitute an effective treatment for osteoarthritis. The chromosomal localization is relevant because it is known that osteoarthritis has a higher frequency in women. It is possible, therefore, that the truncated protein may be nullifying the effect of adlcan, if any, in males.

The NOV61 disclosed in this invention is predicted to be expressed in at least the following tissues: muscle, lymph. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV61 is provided in Example 2.

The nucleic acids and proteins of NOV61 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: osteoarthritis, asthma, allergy, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, cancer, tissue degeneration as well as other diseases, disorders and conditions.. The NOV61 nucleic acid encoding the adlcan-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding an adlcan-like protein includes the nucleic acid whose sequence is provided in Table 61A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 61A while still encoding a protein that maintains its adlcan-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids

whose sequences are complementary to the sequence of Table 61A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 7% of the bases may be so changed.

The novel protein of the invention includes the adican-like protein whose sequence is provided in Table 61B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 61B while still encoding a protein that maintains its adican-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 13% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV62

NOV62 has homology to neuropsin, an extracellular matrix serine protease. The disclosed NOV62 (alternatively referred to herein as CG56781-01) includes the 834 nucleotide sequence (SEQ ID NO:221) shown in Table 62A. A NOV62 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 31-33 and ends with a TGA codon at nucleotides 808-810. The disclosed NOV62 maps to human chromosome 19.

Table 62A. NOV62 Nucleotide Sequence (SEQ ID NO:221)

```
AGTCTTGCCTTCTTTTGAGCCTAAGTCATGAGTTGGATGTTCCCTCAGAGATCTCCTGAGT
GGAGTAAATAAATACTCCACTGGGACTGGATGGATTGGCTGGCTGTCGTGTTTGTCTTC
CGTTTGCTGGTCTACATGGTGGCAGCAGACGCTGGAAAAGATGAGCAGAAAGAGTTT
GAGTGCAACAGTAGACAGCCCGTTGCAAAAATGTGTGTTTGTGACTTCTTCCCCATT
TCCAAGTCAGACTTTGGGCCTTACAACGTGATAATGGTCTCCACACCTTCACTTCTGGTG
GTTTTACATGTAGCCTATCATGAGGGTAGAGAGAAAAGGCACAGAAAGAACTCTATGTC
AGCCCAGGTACAATGGATGGGGCCTATGGTACGCTTATCTTATCAGCCTCATTGTTAAA
ACTGGTTTTGAAATTGGCTTCTTGTGTTTTATTTATAAGCTATATGATGGCTTTAGTGTT
CCCTACCTTATAAAGTGTGATTTGAAGCCTTGTCCCAACACTGTGGACTGCTTCATCTCC
```

AAACCCACTGAGAAGACGATCTTCATCCTCTTCTTGGTCATCACCTCATGCTTGTGTATT
GTGTGAATTTCATTGAAGTGAAGTCTTCTCAAGTGCTTTATTAAGTGCTGTCTC
CAAAAATATTTAAAAAACCTCAAGTCTCAGTGTGTGAGTGCCACAGCCTCAGATATGT
TGATGTG

The NOV62 polypeptide (SEQ ID NO:222) encoded by SEQ ID NO:221 is 259 amino acids in length and is presented using the one-letter amino acid code in Table 62B. The Psort profile for NOV62 predicts that this sequence has a signal peptide and is likely to be localized to the endoplasmic reticulum (membrane) at the plasma membrane with a certainty of 0.5500. In alternative embodiments, a NOV62 polypeptide is located to lysosomes with a certainty of 0.2353, or to the endoplasmic reticulum (lumen) with a certainty of 0.1000. The Signal P predicts a likely cleavage site for a NOV62 peptide is between positions 28 and 29, *i.e.*, at the dash in the sequence TRA-QG.

Table 62B. NOV62 Polypeptide Sequence (SEQ ID NO:222)

MGRPPPCAIQPWILLLLFMGAWAGVTRAQGSRSRKQASKPHSQPWQAALFQGERLICGG
VLVGDRWVLTAAHCKKQKYSVRLGDHSLQSRDQPEQEIQVAQSIQHPCYNNNSNPEDHSHD
IMLIRLQNSANLGDKVKPVQLANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVKIYS
QNK CERAYPGKITEGMVCAGSSNGADTCQGDSSGGLVCEGTLAGIVSGGSEPVFRPRRPA
VYTNVFDYLEWIESTMEKN

A BLAST analysis of NOV62 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV62 had high homology to other proteins as shown in Table 62C.

Table 62C. BLASTX results from PatP database for NOV62

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAW10694 Human recombinant neuropsin	1247	8.9e-127
patp:AAW12393 Mouse neuropsin protein	1247	8.9e-127
patp:AAB21311 Human neuropsin - <i>Homo sapiens</i> , 275 aa.	972	1.2e-97
patp:AA41744 Human PRO322 protein sequence	965	6.8e-97
patp:AA32852 Human serine protease protein sequence	965	6.8e-97

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 747 of 805 bases (92%) identical to a gb:GENBANK-ID:E12348|acc:E12348.1 mRNA from Mus sp. The full amino acid sequence of the protein of the invention was found to have 227 of 257 amino acid residues (88%) identical to, and 238 of 257 amino acid residues (92%) similar to, the 260 amino acid residue ptrn:SWISSNEW-

ACC:Q61955 protein from *Mus musculus* (Mouse) (NEUROPSIN PRECURSOR (EC 3.4.21.) (NP)). NOV62 also has homology to the other proteins shown in the BLASTP data in Table 62D.

Table 62D. NOV62 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 6679487 ref NP_032966.1 (NM_008940)	protease, serine, 19 (neuropsin); Brain Serine protease 1 [<i>Mus musculus</i>]	260	227/257 (88)	238/257 (92)	e-131
gi 6093538 sp O88780 NRPN_RAT	NEUROPSIN PRECURSOR (NP) (BRAIN SERINE PROTEASE 1)	260	216/258 (83)	232/258 (89)	e-126
gi 4699764 pdb 1NPM A	Chain A, Neuropsin, A Serine Protease Expressed In The Limbic System Of Mouse Brain	225	197/223 (88)	206/223 (92)	e-112
gi 6005844 ref NP_009127.1 (NM_007196)	kallikrein 8 (neuropsin/ovasin); protease, serine, 19 (neuropsin/ovasin) [<i>Homo sapiens</i>]	260	172/253 (67)	207/253 (80)	8e-99
gi 16162680 ref XP_057595.1 (XM_057595)	hypothetical protein XP_057595 [<i>Homo sapiens</i>]	260	169/253 (66)	203/253 (79)	3e-96

5

This BLASTP data is displayed graphically in the ClustalW in Table 62E. A multiple sequence alignment is given, with the NOV62 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 62D.

10

Table 62E. ClustalW Alignment of NOV62					
NOV62	(SEQ ID NO:222)				
gi 6679487	(SEQ ID NO:609)				
gi 6093538	(SEQ ID NO:610)				
gi 4699764	(SEQ ID NO:611)				
gi 6005844	(SEQ ID NO:612)				
gi 16162680	(SEQ ID NO:613)				
		10 20 30 40 50 60			
NOV62	1	MGRPPPCAIQPWLLLLFMGAWAGVTRAQGSRSRKQASKPHSQPWQAALFOGERLICGG	60		
gi 6679487	1	MGRPPPCAIQPWLLLLFMGAWAGLTRAQGSKILEGREGIPHSQPWQAALFOGERLICGG	60		
gi 6093538	1	MGRPPPCAIQPWLLFLLMGAWAGLTRAQGSKILEGCECKPHSQPWQIALFOGERLVCGG	60		
gi 4699764	1	-----ILEGREGIPHSQPWQAALFOGERLICGG	28		
gi 6005844	1	MGRPRPRAAKTWMLLLLLGAWAGHSRAQEDKVLGGHECOPHSQPWQAALFOGQQLICGG	60		
gi 16162680	1	MGRPRPRAAKTWMLLLLLGAWAGHSRAQEDKVLGGHECOPHSQPWQAALFOGQQLICGG	60		
		70 80 90 100 110 120			
NOV62	61	VLVGDRWVLTAAHCKKQKYSVRLGDHSLQSRDQPEQEIQVAQSIQHPCYNNNSNPEDHSHD	120		
gi 6679487	61	VLVGDRWVLTAAHCKKQKYSVRLGDHSLQSRDQPEQEIQVAQSIQHPCYNNNSNPEDHSHD	120		
gi 6093538	61	VLVGDRWVLTAAHCKKQKYSVRLGDHSLQKRDEPEQEIQVARSIQHPCENSSNPEDHSHD	120		
gi 4699764	29	VLVGDRWVLTAAHCKKQKYSVRLGDHSLQSRDQPEQEIQVAQSIQHPCYNNNSNPEDHSHD	88		

gi 6005844	61	VLVGNWVLTAAHCKKPKYIVRLGDHSLQNKDGPEQEIPVVSQIPHPCYNSSDVEDHNHD	120
gi 16162680	61	VLVGNWVLTAAHCKKPKYIVRLGDHSLQNKDGPEQEIPVVSQIPHPCYNSSDVEDHNHD	120
	130.....140.....150.....160.....170.....180.....	
NOV62	121	IMLIRLQNSANLGDKVKPVQLANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVKIYS	180
gi 6679487	121	IMLIRLQNSANLGDKVKPVQLANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVKIYS	180
gi 6093538	121	IMLIRLQNSANLGDKVKPIELANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVKIYS	180
gi 4699764	89	IMLIRLQNSANLGDKVKPVQLANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVKIYS	148
gi 6005844	121	IMLLQLRDOASLGSKVKPISLADHCTQPGQKCTVSGWGTVTSPRENFPDNLNCAEVKIEP	180
gi 16162680	121	IMLLQLRDOASLGSKVKPISLADHCTQPGQKCTVSGWGTVTSPRENFPDNLNCAEVKIEP	180
	190.....200.....210.....220.....230.....240.....	
NOV62	181	QNK CERAYPGKITEGMVCAGSSNGADTCQGDSSGGLVCEGTLGIVSGGSEPVFRPRRPA	240
gi 6679487	181	QNK CERAYPGKITEGMVCAGSSNGADTCQGDSSGGLVCDGMLQGITSWGS DPCGKPEKPG	240
gi 6093538	181	QNK CERAYPGKITEGMVCAGSSNGADTCQGDSSGGLVONGVLQGITWGS DPCGKPEKPG	240
gi 4699764	149	QNK CERAYPGKITEGMVCAGSSNGADTCQGDSSGGLVCDGMLQGITSWGS DPCGKPEKPG	208
gi 6005844	181	QKKCEDAYPGQITDGMVCAGSSK GADTCQGDSSGGLVCDGALQGITSWGS DPCGRSDKPG	240
gi 16162680	181	QKKCEDAYPGQITDGMVCAGSSK GADTCQGDSSGGLVCDGALPGHHITGLRPPCGRSDKPG	240
	250.....260.....	
NOV62	241	VYTNVFDYLEWIESTMEKN-	259
gi 6679487	241	VYTKICRYTTWIKKTM DNRD	260
gi 6093538	241	VYTKICRYTNWIKKTMGKRD	260
gi 4699764	209	VYTKICRYTTWIKKTM D---	225
gi 6005844	241	VYTNICRYLDWIKKIIGSKG	260
gi 16162680	241	VYTNICRYLDWIKKIIGSKG	260

Table 62F lists the domain description from DOMAIN analysis results against NOV62. This indicates that the NOV62 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 62F. Domain Analysis of NOV62				
gnl Smart smart00020, Tryp_SPc, Trypsin-like serine protease; Many of these are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis to generate their active forms. A few, however, are active as single chain molecules, and others are inactive due to substitutions of the catalytic triad residues. SEQ ID NO:812				
CD-Length = 230 residues, 98.3% aligned Score = 237 bits (604), Expect = 7e-64				
NOV62:	36	GQASKPHSQPWQAALF-QGERLICGGVLVGRWVLTAAHC----	KKQKYSVRLGDHSLQS	90
Sbjct:	5	GSEANIGSFPWQVSLQYRGRHFCGGLISPRWVLTAAHC	VRLG H L S	64
NOV62:	91	RDQPEQEIQVAQSIQHPCYNNSNPEDHSHDIMLIRLQNSANLGDKVKPVQLA--NLCPKV		148
Sbjct:	65	GE-ETQTVKVSQVIVHPNYN---PSTYDNDIALKLSEPVTLSDTVRPICLPSSGYNVPA		120
NOV62:	149	GQKCIISGWGTVTSPQENFPNTLNCAEVKIYSQNK CERAYPG--KITEGMVCAGSSN-GA		205
Sbjct:	121	GTTCTVSGWGRTSESSGSLPDTLQEVNVPVIVSNATCRRAYSGGPAITDNMLCAGGLEGGK		180
NOV62:	206	DTCQGDSSGGLVCE---GTLGIVSGGSEPVFRPRRPAVYTNVFDYLEWI		252

	D	CQGD	SGG	PLVC	L	GIVS	GS	RP	+P	VYT	V	YL+WI	
Subjct:	181	DACQGD	SGG	PLVCND	PRW	LVG	IVSWG	SYGC	ARNK	PGVY	TRV	SSYLD	WI 230

Neuropsin appears to act as a regulatory molecule in the early phase of LTP via its proteolytic function on extracellular matrix rather than affecting NMDA receptor-mediated calcium increase. The behavioral and electrographical abnormalities associated with seizures in epileptic (kindled) mice correspond with those of human epilepsy. In kindled mice, neuropsin was markedly increased in the hippocampus and cerebral cortices. A single intraventricular injection of monoclonal antibodies specific to neuropsin reduced or eliminated the epileptic pattern noted on electroencephalograms and, as a result markedly inhibited the progression of kindling.

Therefore, neuropsin appears to be a key protein controlling pathogenic events in the hippocampus, and thus neuropsin inhibitors might be useful for treatment of epilepsy. Neuropsin has two isoforms, which have been reported to be involved in hippocampal plasticity. The amino acid sequences of the two types of human neuropsin were identical, except that type 2 carried an insert of 45 amino acids at the C-terminus of the leader sequence. The essential three amino acids in the active site triad, His, Asp, and Ser, and the single putative N-glycosylation site were conserved in human and mouse neuropsin. Sequence analysis of the 946 bp genomic DNA spanning the region encoding the insertion sequence revealed that two isoforms were generated in human brain by alternative splicing. However, the mouse genomic sequence did not conserve the 3' acceptor consensus sequence at the corresponding position, suggesting that type 2 neuropsin was a species-specific splice variant. When the open reading frames of human neuropsin were expressed in insect cells, both types of neuropsin were detected in the conditioned media by western blot analysis using anti-human neuropsin serum.

Northern blot hybridization and reverse transcription-polymerase chain reaction showed predominant expression of type 1 neuropsin in pancreas. Type 2 neuropsin was preferentially expressed in human adult brain and hippocampus, although both types were expressed in fetal brain and placenta in comparable amounts. Dot blot hybridization showed that neuropsin was expressed in various regions of adult brain, including the hippocampus and cerebral cortex, and also in various fetal tissues. These results suggest that human type 2 neuropsin may be important to the adult brain plasticity, although both types may be necessary for the development of the nervous system.

The disclosed NOV62 is predicted to be expressed in at least the following tissues: brain. This information was derived by determining the tissue sources of the sequences that

were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV62 is provided in Example 2.

5 The nucleic acids and proteins of NOV62 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for treatment of patients suffering from: Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral
10 disorders, addiction, anxiety, pain, neuroprotection, osteoarthritis, and other diseases, disorders and conditions of the like. The NOV62 nucleic acid encoding the neuropsin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a neuropsin precursor-like protein includes the
15 nucleic acid whose sequence is provided in Table 62A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 62A while still encoding a protein that maintains its neuropsin precursor-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to
20 those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar
25 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 8% of the residues may be so changed.

30 The novel protein of the invention includes the neuropsin precursor-like protein whose sequence is provided in Table 62B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 62B while still encoding a protein that maintains its neuropsin precursor-like activities and physiological

functions, or a functional fragment thereof. In the mutant or variant protein, up to about 12% of the bases may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV63

NOV63 has homology to the WNT family of proteins. The Wnt gene family consists of at least 15 structurally related genes that encode secreted extracellular signaling factors. WNT proteins function in a range of critical developmental processes in both vertebrates and invertebrates and are implicated in regulation of cell growth and differentiation in certain adult mammalian tissues, including the mammary gland. The disclosed NOV63 (alternatively referred to herein as CG56054-02) includes the 1128 nucleotide sequence (SEQ ID NO:) shown in Table 63A. A NOV63 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 31-33 and ends with a stop codon at nucleotides 1102-1104. The disclosed NOV63 maps to human chromosome 1.

Table 63A. NOV63 Nucleotide Sequence (SEQ ID NO:223)

```
TCCTCCCCGAGCTTCTCGCTGAATCCGAGGGGGCTGAGAGGATGGCCACCACCGGGAC
CCCAACGGCCGACCGAGGGCAGCGACGCCACAGATGACCCGGCCGCCCGCTTCCAGGT
GCAGAAGCACTCGTGGGACGGGCTCCGGAGCATCATCCACGGCAGCCGCAAGTACTCGGG
CCTCATTGTCAACAAGGCGCCCCACGACTTCCAGTTTGTGTCAGAAGACGGATGAGTCTGG
GCCCCACTCCACCGCCTCTACTACCTGGGAATGCCATATGGCAGCCGAGAGAAGTCCCT
CCTCTACTCTGAGATTCCCAAGAAGGTCCGGAAGAGGCTCTGCTGCTCCTGTCTCTGGAA
GCAGATGCTGGATCATTTCAGGCCACGCCCCACCATGGGGTCTACTCTCGGGAGGAGGA
GCTGCTGAGGGAGCGGAAACGCCTGGGGGTCTTCGGCATCACCTCCTACGACTTCCACAG
CGAGAGTGGCCTCTCCTCTTCCAGGCCAGCAACAGCCTCTTCCACTGCCGCGACGGCGG
CAAGAACGGCTTCATGGTGTCCCTATGAAACCGCTGGAAATCAAGACCCAGTGTCTCAGG
GCCCCGGATGGACCCCAAAATCTGCCCTGCCGACCCTGCCTTCTTCTCCTTCATCAATAA
CAGCGACCTGTGGGTGGCCAACATCGAGACAGGCGAGGAGCGGCGGCTGACCTTCTGCCA
CCAAGGTTTATCCAATGTCTGGATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCAT
ACAGGAAGAGTTCGACCGCTTCACTGGGTACTGGTGGTGCCCCACAGCCTCCTGGGAAGG
TTCAGAGGGCCTCAAGACGCTGCGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGA
GGTCATTACGTCCCCTCTCCTGCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCC
CAGGACAGGTAGCAAGAATCCCAAGATTGCCCTGAAACTGGCTGAGTTCAGACTGACAG
CCAGGGCAAGATCGTCTCGACCCAGGAGAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTT
CCCGAAGGTGGAGTACATCGCCAGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGC
CATGTTCTTGGACCGGCCCCAGCAGTGGCTCCAGCTCGTCTCCTCCCCCGGCCCTGTT
CATCCCGAGCACAGAGAATGAGGAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAA
TGTCCAGCCGTATGTGGTGTACGAGGAGGTACCAACGTCTGGATCAATGTTTCATGACAT
CTTCTATCCCTTCCCCAATCAGAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGA
ATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCACCGCGGTTTAAATCCAGGGCTA
CGATTGGAGTGAGCCCTTCAGCCCCGGGAAGATGAATTTAAGTGCCCCATTAAGGAAGA
GATTGCTCTGACCAGCGGTGAATGGGAGGTTTGGCGAGGCACGCTCCAAGATCTGGGT
CAATGAGGAGACCAAGCTGGTGTACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCA
```


CCTCTACGTGGTCAGCTATGAGGCGGCCGCGAGATCGTACGCCTCACCAGCCCGGCTT
CTCCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTTTCGTACGCCACTACAGCAGCGT
GAGCAGCCCGCCCTGCGTGACGCTCTACAAGCTGAGCGGCCCGACGACGACCCCTGCA
CAAGCAGCCCGCTTCTGGGCTAGCATGATGGAGGCAGCCAGCTGCCCCCGGATTATGT
TCCTCCAGAGATCTTCCATTTCCACACGCGCTCGGATGTGCGGCTCTACGGCATGATCTA
CAAGCCCCACGCCTTGACGCCAGAGAAGAAGCACCCACCGTCTCTTTGTATATGGAGG
CCCCCAGGTGCAGCTGGTGAATAACTCCTTCAAAGGCATCAAGTACTTGCGGCTCAACAC
ACTGGCCTCCCTGGGCTACGCCGTGGTTGTGATTGACGGCAGGGGCTCCTGTCAGCGAGG
GCTTCGGTTCGAAGGGGCCCTGAAAAACCAATGGGCCAGGTGGAGATCGAGGACCAGGT
GGAGGGCCTGCAGTTCGTGGCCGAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCAT
CCATGGCTGGTCTACGGGGGCTTCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGT
GTTCAAGGTGGCCATCGCGGGTCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTA
CACTGAGCGCTACATGGACGTCCTGAGAACAACAGCACGGCTATGAGGCGGGTTCGGT
GGCCCTGCACGTGGAGAAGCTGCCCAATGAACCAACCGTTGCTTATCCTCCACGGCTT
CCTGGACGAAAACGTGCACTTTTTCCACACAACTTCTCGTCTCCCACTGATCCGAGC
AGGGAAACCTTACCAGCTCCAGATCTACCCCAACGAGAGACACAGTATTCGCTGCCCGGA
GTCGGGCGAGCACTATGAAGTCACGTTGCTGCACCTTCTACAGGAATACCTCTGAGCCTG
CCCACCGGAGCCGCCACAT

The NOV63 polypeptide (SEQ ID NO:224) encoded by SEQ ID NO:223 is 357 amino acids in length and is presented using the one-letter amino acid code in Table 63B. The Psort profile for NOV63 predicts that this sequence has a signal peptide and is likely to be exported from the cell with a certainty of 0.3700. In alternative embodiments, a NOV63 polypeptide is located to lysosomes with a certainty of 0.1000, or to the endoplasmic reticulum (membrane) with a certainty of 0.1000. The Signal P predicts a likely cleavage site for a NOV63 peptide is between positions 18 and 19, *i.e.*, at the dash in the sequence ALG-SY.

Table 63B. NOV63 Polypeptide Sequence (SEQ ID NO224)

MATTGTPTADRGDAAATDDPAARFQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQ
KTDESGPHSHRLYYLGMPYGSRENSLLYSEIPKKVRKEALLLSWKQMLDHFQATPHHGV
YSREEELLREKRRLGVFGITSYDFHSESGFLFQASNSLFHCRDGGKNGFMVSPMKPLEI
KTQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGEERRLTFCHQGLSNVLDPPKSAG
VATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRIlyEEVDESEVEVIHVPSPALEERKT
DSYRYPRTGSKNPKIALKLAEFQTD SQGKIVSTQEKELVQPFSSLPKVEYIARAGWTRD
GKYAWAMFLDRPQQLVLLPPLFIPSTENEEQRLASARAVPRNVQPYVVVEVTNVW
INVHDI FYPPQSEGEDEL CFLRANECKTG FCHLYKVTA VLKSQGYDWSEPFSPGEDEFK
CPIKEEIALTSGEWEVLARHGSKIWNNEETKL VYFQGT KDTPLEHHLYVVS YEAGEIVR
LTPPGF SHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPD DPLHKQPRFWASMMEAS
CPPDYVPPEIFHFHTRSDVRLYGM IYKPHALQPEKKHPTVLFVYGGPQVQLVNNSFKGIK
YLRLNTLASLGYAVVVIDGRGSCQRGLRFEGALKNQMGQVEIEDQVEGLQFVAEKYGFID
LSRVAIHGWSYGGFLSLMGLIHK PQVFKA IAGAPVTVMAYDTGYTERYMDVPENNQHG
YEAGSVALHVEKLPNEPNRLLILHGF LDENVHFFHTNFLVSQ LIRAGKPYQLQIYPNERH
SIRCPESGEHYEVTLLHFLQEYL

A BLAST analysis of NOV63 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV63 had high homology to other proteins as shown in Table 63C.

Table 63C. BLASTX results from PatP database for NOV63

Smallest

Sequences producing High-scoring Segment Pairs:	High Score	Sum Probability P(N)
patp:AAY81693 Human Wnt-6 protein sequence - <i>Homo sapiens</i>	411	1.0e-61
patp:AAB49769 Amyloid-beta protein agglutination regulator	411	1.0e-61
patp:AAB88439 Human membrane or secretory protein	411	1.0e-61
patp:AAB19786 Human Wnt-1 protein involved in kidney	630	2.2e-61
patp:AAY94319 Murine Wnt-10A protein - <i>Mus musculus</i> , 417 aa.	398	2.7e-61

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 633 of 795 bases (79%) identical to a gb:GENBANK-ID:AF031168|acc:AF031168.1 mRNA from *Gallus gallus* (*Gallus gallus* Wnt-14 protein (Wnt-14) mRNA). The full amino acid sequence of the protein of the invention was found to have 287 of 349 amino acid residues (82%) identical to, and 319 of 349 amino acid residues (91%) similar to, the 354 amino acid residue ptnr:SWISSPROT-ACC:O42280 protein from *Gallus gallus* (Chicken) (WNT-14 PROTEIN PRECURSOR). NOV63 also has homology to the other proteins shown in the BLASTP data in Table 63D.

10

Table 63D. NOV63 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 15082261 ref NP_003386.1 (NM_003395)	wingless-type MMTV integration site family, member 14 [<i>Homo sapiens</i>]	365	339/356 (95)	343/356 (96)	0.0
gi 3915306 sp O42280 WN14_CHICK	WNT-14 PROTEIN PRECURSOR	354	283/333 (84)	310/333 (92)	e-156
gi 16303264 dbj BAB70499.1 (AB063483)	WNT14B [<i>Homo sapiens</i>]	357	215/352 (61)	263/352 (74)	e-107
gi 17017976 ref NP_003387.1 (NM_003396)	wingless-type MMTV integration site family, member 15 precursor [<i>Homo sapiens</i>]	357	215/352 (61)	263/354 (74)	e-107
gi 18181917 dbj BAB83866.1 (AB073819)	Wnt14b [<i>Mus musculus</i>]	359	216/354 (61)	264/354 (74)	e-106

This BLASTP data is displayed graphically in the ClustalW in Table 63E. A multiple sequence alignment is given, with the disclosed NOV63 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 63D.

15

Table 63E. ClustalW Alignment of NOV63	
NOV63	(SEQ ID NO:224)
gi 15082261	(SEQ ID NO:614)
gi 3915306	(SEQ ID NO:615)

gi 16303264	(SEQ ID NO:616)
gi 17017976	(SEQ ID NO:617)
gi 18181917	(SEQ ID NO:618)

102030405060

NOV63

1

1

1

1

1

1

---MAPLCGYFLLCS---LKQAL--G-SYPIWLTGSEPLTILP-LTLEP-EAAAQAHY48

MLDGSPLARWLAAAFGLTLLAAL--RPSAAYFGLTGSEPLTILP-LTLEP-EAAAQAHY56

---MALLRALLGLLA-----CTP--RPSAAYFGLTGNEALTILP-LTSEMEEAAVKAHY48

---MRPPPALALAGLC---LLALP--AAAASYFGLTGREVLTPFPGLGTAAAPAQGGAHL52

---MRPPPALALAGLC---LLALP--AAAASYFGLTGREVLTPFPGLGTAAAPAQGGAHL52

---MRPPALALAAALC---LLVLPAAAAAAYFGLTGREVLTPFPGLGTAAAPAQAGGAHL54

708090100110120

NOV63

49

57

49

53

53

55

KACDRLKLERKORMCRRDPGVAETLVEAVSMSALECQFQFRERWNCILEGRYRASLLK108

KACDRLKLERKORMCRRDPGVAETLVEAVSMSALECQFQFRERWNCILEGRYRASLLK116

KVCDRLKLEKKORMCRRDPGGAETLMEAI SMSALECQYQFRERWNCILEGRYRASLLK108

KQCDLLKLSRRQKQCRREPGLAETLRDAHLGLLECQFQFRHERWNCLEG--RTGLLK110

KQCDLLKLSRRQKQCRREPGLAETLRDAHLGLLECQFQFRHERWNCLEG--RMGLLK110

KQCDLLKLSRRQKQCRREPGLAETLRDAHLGLLECQFQFRERWNCLEG--RTGLLQ112

130140150160170180

NOV63

109

117

109

111

111

113

RGFKETAFLYAISSAGLTHALAKACSAGRMERCTCDEAPDLENREAWQWGC CGDNLKYSS168

RGFKETAFLYAISSAGLTHALAKACSAGRMERCTCDEAPDLENREAWQWGC CGDNLKYSS176

RGFKETAFLYAISSAGLTHAMAKACSAGRMERCTCDEAPDLENREAWQWGC CGDNLKYSN168

RGFKETAFLYAVSSAALHTHLARACSAGRMERCTCDDSPGLESRCQAWQWGC CGDNLKYST170

RGFKETAFLYAVSSAALHTHLARACSAGRMERCTCDDSPGLESRCQAWQWGC CGDNLKYST170

RGFKETAFLYAVSSAALHTHALARACSAGRMERCTCDDSPGLESRCQAWQWGC CGDNLKYST172

190200210220230240

NOV63

169

177

169

171

171

173

KFVKEFLG-RRSSKDLRARVDFHNNLVGVKVIKAGVETTCCKCHGVSGSCTVRTCWRQLAP227

KFVKEFLG-RRSSKDLRARVDFHNNLVGVKVIKAGVETTCCKCHGVSGSCTVRTCWRQLAP235

KFVKEFLG-RKPNKDLRARVDFHNNLVGVKVIKAGVETTCCKCHGVSGSCTVRTCWRQLSP227

KFLSNFLGSKRGKDLRARADAHNTHVGITKAVKSGLRITTCCKCHGVSGSCAVRTCWKQLSP230

KFLSNFLGSKRGKDLRARADAHNTHVGITKAVKSGLRITTCCKCHGVSGSCAVRTCWKQLSP230

KFLSNFLGPKRGSKDLRARADAHNTHVGITKAVKSGLRITTCCKCHGVSGSCAVRTCWKQLSP232

250260270280290300

NOV63

228

236

228

231

231

233

FHEVGKHLKHXYETALKVGSTTNEAAGEAGAI SPPRGRASCAGGSDPLPRTPELVHLLDS287

FHEVGKHLKHXYETALKVGSTTNEAAGEAGAI SPPRGRASCAGGSDPLPRTPELVHLLDS295

FHEIGKQLKQKYETSLKVGSTTNEATGE-GDISPPK--KSIPGHSDQIPRTIDLVYIDS284

FRETGQVLKRLRYDSAVKVSSATNEALGRLELWAPAR---QGSLTGKLAPRSGDLVYMEDS287

FRETGQVLKRLRYDSAVKVSSATNEALGRLELWAPAR---QGSLTGKLAPRSGDLVYMEDS287

FRETGQVLKRLRYDTAVKVSSATNEALGRLELWAPAK---PGGPAKLAPRPGDLVYMEDS289

310320330340350360

NOV63

288

296

285

288

288

290

PSFCLAGRFSPGTAGRRCHREKNCE SICCGRGHNTQSRVVTRPCCQVWRCCYVECRQCT347

PSFCLAGRFSPGTAGRRCHREKNCE SICCGRGHNTQSRVVTRPCCQVWRCCYVECRQCT355

PSFCLMSRYSPGTSGRKCYKDKNCDS ICCGRGHNTQSRVVTRPCCQVWRCCYVECKQCT344

PSFCRPSKYSPGTAGRVCSREASCSSLCGRGYDTQSRVAFSCHCQVQWCCYVECCQCV347

PSFCRPSKYSPGTAGRVCSREASCSSLCGRGYDTQSRVAFSCHCQVQWCCYVECCQCV347

PSFCRPSKYSPGTAGRVCSRSSCSSLCGRGYDTQSRMVVFSCHCQVQWCCYVECCQCA349

370

NOV63

348

356

345

348

348

350

QREEVYTCKG357

QREEVYTCKG365

QREEVYTCKD354

QEELVYTCKH357

QEELVYTCKH357

QEELVYTCKR359

Table 63F lists the domain description from DOMAIN analysis results against NOV63. This indicates that the NOV63 sequence has properties similar to those of other proteins known to contain this domain.

Table 63F. Domain Analysis of NOV63					
gnl Pfam pfam00110, wnt, wnt family. SEQ ID NO:862					
CD-Length = 313 residues, 99.7% aligned					
Score = 283 bits (725), Expect = 9e-78					
NOV63:	51	CDRLK-LERKQRRMCRDPGVAETLVEAVSMSALECQFQFRFERWNTLEGRYRASL---	106		
Sbjct:	2	C L L +QR++CRR+P V ++ E ++ ECQ QFR RWNC+ R R	61		
NOV63:	107	LKRGFKETAFLYAIISSAGLTHALAKACSAGRMERCTCDE-APDLENREAWQWGGCGDNLK	165		
Sbjct:	62	LK+G +ETAF+YAISSAG+ HA+ +ACS G +E C CD + +WQWGGC DN++	121		
NOV63:	166	LKKGTRETAFFVYAISSAGVAHAVTRACSEGELESCGCDYKKGPGGPQGSWQWGGCSDNVE	224		
Sbjct:	122	+ +F +EF+ R +D R+ ++ HNN G K +K+ + CKCHGVSGSC++TCW	181		
NOV63:	225	FGIRFSREFVDARERERDARSLMNLHNEAGRKAVKSHMRRECKCHGVSGSCSMKTCWLS	283		
Sbjct:	182	LAPFHEVGKHLKHKYETALKV-GSTTNEAAGEAGAI SPPRGRASGAGSDPLRTPPELVH	234		
NOV63:	284	L F VG LK KY+ A++V + G A + R SD LV+	337		
Sbjct:	235	L+DSP +C S GT GR C CE +CCGRG+NTQ T C C+ W	294		
NOV63:	338	LEDSPDYCERDRSTGSLGTQGRVCNKTSGKLDGCELLCCGRGYNTQQVERTEKCNCKFHW	356		
Sbjct:	295	CCYVECRQCTQREEVYTCK CCYV+C +C + EV+TCK	313		

5

Wnt-1 (previously known as int-1) is a proto-oncogene induced by the integration of the mouse mammary tumor virus. It is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). The sequence of wnt-1 is highly conserved in mammals, fish, and amphibians.

10 Wnt-1 is a member of a large family of related proteins that are all thought to be developmental regulators. These proteins are known as wnt-2 (also known as irp), wnt-3, up to wnt-15. At least four members of this family are present in Drosophila. One of them, wingless (wg), is implicated in segmentation polarity. All these proteins share the following features

15 characteristics of secretory proteins, a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters.

The NOV63 disclosed in this invention is predicted to be expressed in at least the following tissues: brain This information was derived by determining the tissue sources of the

sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV63 is provided in Example 2.

5 The nucleic acids and proteins of NOV63 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for treatment of patients suffering from: Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral
10 disorders, addiction, anxiety, pain, neuroprotection, osteoarthritis, and other diseases, disorders and conditions of the like. The NOV63 nucleic acid encoding the WNT-14 precursor-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a WNT-14 precursor-like
15 protein includes the nucleic acid whose sequence is provided in Table 63A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 63A while still encoding a protein that maintains its WNT-14 precursor-like activities and physiological functions, or a fragment
20 of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence indicated in Table 63A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

 The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar
25 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 21% of the bases may be so changed.

30 The novel protein of the invention includes the WNT-14 precursor-like protein whose sequence is provided in Table 63B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 63B while still encoding a protein that maintains its WNT-14 precursor-like activities and physiological

functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV64

NOV64 has homology to dipeptidyl peptidase. The disclosed NOV64 (alternatively referred to herein as CG56884-01) includes the 2660 nucleotide sequence (SEQ ID NO:225) shown in Table 64A. A NOV64 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 44-46 and ends with a stop codon at nucleotides 2633-2635. The disclosed NOV64 maps to human chromosome 17.

Table 64A. NOV64 Nucleotide Sequence (SEQ ID NO:225)

```
TCCTCCCCGCGAGCTTCTCGCTGAATTCCGAGGGGGCTGAGAGGATGGCCACCACCGGGAC
CCCAACGGCCGACCGAGGCGACGCGAGCCGCCACAGATGACCCGGCCGCGCCGCTTCCAGGT
GCAGAAGCACTCGTGGGACGGGCTCCGAGGAGCATCATCCACGGCAGCCGCAAGTACTCGGG
CCTCATTGTCAACAAGGCGCCCCACGACTTCCAGTTTGTGTCAGAAGACGGATGAGTCTGG
GCCCCACTCCCACCGCCTCTACTACCTGGGAATGCCATATGGCAGCCGAGAGAACTCCCT
CCTCTACTCTGAGATTCCCAAGAAGGTCCGGAAAGAGGCTCTGCTGCTCCTGTCTGGAA
GCAGATGCTGGATCATTTCAGGCCACGCCCCACCATGGGGTCTACTCTCGGGAGGAGGA
GCTGCTGAGGAGCGGAAACGCTGGGGTCTTCGGCATCACCTCCTACGACTTCCACAG
CGAGAGTGGCCTCTTCTCTTCCAGGCCAGCAACAGCCTCTTCCACTGCCGCGACGGCGG
CAAGAACGGCTTCATGGTGTCCCTATGAAACCGCTGGAATCAAGACCCAGTGCTCAGG
GCCCCGGATGGACCCAAATCTGCCCTGCCGACCCTGCCTTCTTCTCTCATCAATAA
CAGCGACCTGTGGGTGGCCAAACATCGAGACAGGCGAGGAGCGCGGCTGACCTTCTGCCA
CCAAGGTTTATCCAATGTCTGGATGACCCCAAGTCTGCGGGTGTGGCCACCTTCGTCTAT
ACAGGAAGAGTTCGACCGCTTCACTGGGTACTGGTGGTGCCCCACAGCCTCTGGGAAGG
TTCAGAGGGCCTCAAGACGCTGCGAATCCTGTATGAGGAAGTCGATGAGTCCGAGGTGGA
GGTCATTACGTCCCCCTCTCTGCGCTAGAAGAAAGGAAGACGGACTCGTATCGGTACCC
CAGGACAGGTAGCAAGAATCCCAAGATTGCCTTGAACTGGCTGAGTTCAGACTGACAG
CCAGGGCAAGATCGTCTCGACCCAGGAGAAAGGAGCTGGTGCAGCCCTTCAGCTCGCTGTT
CCCGAAGGTGGAGTACATCGCCAGGGCCGGGTGGACCCGGGATGGCAAATACGCCTGGGC
CATGTTCTGGACCGGCCCCAGCAGTGGCTCCAGCTCGTCTCTCTCCCCCGGCCCTGTT
CATCCCGAGCACAGAGAATGAGGAGCAGCGGCTAGCCTCTGCCAGAGCTGTCCCCAGGAA
TGTCCAGCCGTATGTGGTGTACGAGGAGGTCACCAACGTCTGGATCAATGTTTCATGACAT
CTTCTATCCCTTCCCCCAATCAGAGGGAGAGGACGAGCTCTGCTTTCTCCGCGCCAATGA
ATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCACCGCGGTTTTAAATCCCAGGGCTA
CGATTGGAGTGAGCCCTTCAGCCCCGGGAAGATGAATTTAAGTGCCCATTAAGGAAGA
GATTGCTCTGACCAGCGGTGAATGGGAGGTTTTGGCGAGGACAGGCTCCAAGATCTGGGT
CAATGAGGAGACCAAGCTGGTGTACTTCCAGGGCACCAAGGACACGCCGCTGGAGCACCA
CCTCTACGTGGTCAGCTATGAGGCGGCGGCGAGATCGTACGCCTCACCACGCCCGGCTT
CTCCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTTTCGTACGCCACTACAGCAGCT
GAGCAGCCGCCCTGCGTGCACGTCTACAAGCTGAGCGGCCCCGACGACGACCCCTGCA
CAAGCAGCCCCGCTTCTGGGCTAGCATGATGGAGGACGCCAGCTGCCCCCGGATTATGT
TCCTCCAGAGATCTTCCATTTCCACACGCGCTCGGATGTGGGCTCTACGGCATGATCTA
CAAGCCCCACGCCTTGACAGCCAGAGAAGAAGACCCCCACCGTCTCTTGTATATGGAGG
CCCCCAGGTGACGTGGTGAATAACTCTTCAAAGGCATCAAGTACTTGGGCTCAACAC
ACTGGCCTCCCTGGGCTACGCCGTGGTGTGATTGACGGCAGGGGCTCTGTGAGCGAGG
GCTTCGGTTTCAAGGGGCCCTGAAAAACCAATGGGCCAGGTGGAGATCGAGGACCAGGT
```

GGAGGGCCTGCAGTTCGTGGCCGAGAAGTATGGCTTCATCGACCTGAGCCGAGTTGCCAT
 CCATGGCTGGTCTACGGGGGCTTCTCTCGCTCATGGGGCTAATCCACAAGCCCCAGGT
 GTTCAAGGTGGCCATCGCGGGTGCCCCGGTCACCGTCTGGATGGCCTACGACACAGGGTA
 CACTGAGCGCTACATGGACGTCCCTGAGAACAACCAGCACGGCTATGAGCGGGTCCGT
 GGCCCTGCAGTGGAGAAGCTGCCAATGAACCCAACCGCTTGCTTATCCTCCACGGCT
 CCTGGACGAAAACGTGCACCTTTTCCACAAAACCTCCTCGTCTCCCAACTGATCCGAGC
 AGGGAAACCTTACCAGCTCCAGATCTACCCCAACGAGAGACACAGTATTCCGTGCCCGGA
 GTCGGGCGAGCACTATGAAGTCACGTTGCTGCACCTTCTACAGGAATACCTCTGAGCCTG
 CCCACCGGAGCCGCCACAT

A NOV64 polypeptide (SEQ ID NO:226) encoded by SEQ ID NO:225 is 863 amino acids in length and is presented using the one-letter amino acid code in Table 64B. The Psort profile for NOV64 predicts that this sequence has no signal peptide and is likely to be localized at peroxisomal microbodies with a certainty of 0.6400. In alternative embodiments, a NOV64 polypeptide is located to lysosomes with a certainty of 0.1000, or to the cytoplasm with a certainty of 0.4500.

Table 64B. NOV64 Polypeptide Sequence (SEQ ID NO:226)

MATGTPTADRGDAAATDDPAARFQVQKHSWDGLRSIIHGSRKYSGLIVNKAPHDFQFVQ
 KTDESGPHSHRLYYLGMPYGSRENSLLYSEIPKKVRKEALLLSWKQMLDHFQATPHHGV
 YSREEELLRERKRLGVFGITSYDFHSEGLFLFQASNSLFHCRDGGKNGFMVSPMKPLEI
 KTQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGEERRLTFC HQGLSNVLD DPKSAG
 VATFVIQEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVIHVPSPALEERT
 DSYRYPRTGSKNPKIALKLAEFQTD SQGKIVSTQEKELVQPFSSLPKVEYIARAGWTRD
 GKYAWAMFLDRPQWLQLVLLPPALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVW
 INVHDIIFYPFQSEGEDEL CFLRANECKTG FCHLYKVTA VLKSQGYDWSEPFSPGEDEFK
 CPIKEEIALTSGEWEVLARHGSKIWNNEETKL VYFQGT KDTPLEHHLYVVS YEAAAGEIVR
 LTPGFSHSCSMSQNFDMFVSHYSSVSTPPCVHVYKLSGPD DDLHKQPRFWASMMEAAS
 CPPDYVPPEIFHFHTRSDVRLYGM IYKPHALQPEKKHPTVLFVYGGPQVQLVMNSFKGIK
 YLRNLTLASLGAVVVIDGRGSCQRGLRFEGALKNQMQVEIEDQVEGLQFVAEKYGFID
 LSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTERYMDVPENNQHG
 YEAGSVALHVEKLPNEPNRLLILHGF LDENVHFFHTN FLVSQ LIRAGKPYQLQIYPNERH
 SIRCPESGEHYEVTLLHFLQEYL

A BLAST analysis of NOV64 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV64 had high homology to other proteins as shown in Table 64C.

Table 64C. BLASTX results from PatP database for NOV64

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability	
		P(N)	
patp:AAB41626 Human ORFX ORF1390 polypeptide sequence	3403	0.0	
patp:AAM38724 Human polypeptide	1987	0.0	
patp:AAM40510 Human polypeptide	1823	0.0	
patp:AAB47187 Human DPP8 - <i>Homo sapiens</i> , 882 aa.	2868	1.5e-298	
patp:AAY90299 Human peptidase, HPEP-16 protein sequence	2547	1.6e-264	

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1601 of 2525 bases (63%) identical to a gb:GENBANK-ID:AF221634|acc:AF221634.1 mRNA from *Homo sapiens* (dipeptidyl peptidase 8 (DPP8) mRNA). The full amino acid sequence of the protein of the invention was found to have 521 of 856 amino acid residues (60%) identical to, and 657 of 856 amino acid residues (76%) similar to, the 882 amino acid residue ptnr:TREMBLNEW-ACC:AAG29766 protein from *Homo sapiens* (Human) (DIPEPTIDYL PEPTIDASE 8). NOV64 also has homology to the other proteins shown in the BLASTP data in Table 64D.

Table 64D. NOV64 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17865311 gb AAL47179.1 AF452102.1 (AF452102)	dipeptidyl peptidase 9 [Homo sapiens]	863	862/863 (99)	862/863 (99)	0.0
gi 17483229 ref XP_035636.2 (XM_035636)	hypothetical protein XP_035636 [Homo sapiens]	689	688/689 (99)	688/689 (99)	0.0
gi 12855335 dbj BAB30295.1 (AK016546)	DIPEPTIDYL PEPTIDASE 8-putative [Mus musculus]	883	519/859 (60)	658/859 (76)	0.0
gi 11095188 gb AAG29766.1 AF221634.1 (AF221634)	dipeptidyl peptidase 8 [Homo sapiens]	882	516/840 (61)	650/840 (76)	0.0
gi 3513303 gb AAC33801.1 (AC005594)	R26984_1 [Homo sapiens]	508	492/543 (90)	495/543 (90)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 64E. A multiple sequence alignment is given, with the NOV64 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 64D.

Table 64E. ClustalW Alignment of NOV64	
NOV64	(SEQ ID NO:226)
gi 17865311	(SEQ ID NO:619)
gi 17483229	(SEQ ID NO:620)
gi 12855335	(SEQ ID NO:621)
gi 11095188	(SEQ ID NO:622)
gi 3513303	(SEQ ID NO:623)
<div> <div>102030405060</div> <div> <div>NOV641</div> <div>gi 178653111</div> <div>gi 174832291</div> </div> <div> <div>-----</div> <div>-----</div> <div>-----</div> </div> <div> <div>MATTGTP</div> <div>MATTGTP</div> <div>MATTGTP</div> </div> <div> <div>ADRGDAA</div> <div>ADRGDAA</div> <div>ADRGDAA</div> </div> <div> <div>TDDE</div> <div>TDDE</div> <div>TDDE</div> </div> <div> <div>-AARFQV</div> <div>-AARFQV</div> <div>-AARFQV</div> </div> <div> <div>QKH</div> <div>QKH</div> <div>QKH</div> </div> <div> <div>SWDGL</div> <div>SWDGL</div> <div>SWDGL</div> </div> <div> <div>LRST</div> <div>LRST</div> <div>LRST</div> </div> <div> <div>THGSR</div> <div>THGSR</div> <div>THGSR</div> </div> <div> <div>KYSGL</div> <div>KYSGL</div> <div>KYSGL</div> </div> <div> <div>---</div> <div>---</div> <div>---</div> </div> <div> <div>48</div> <div>48</div> <div>1</div> </div> </div>	

gi	12855335	1	MAAAMETEQLGVEIFETAECHEGNGESQDRPKLEPFYVERYSWSQLKKLIADTRKYHGYM	60
gi	11095188	1	MAAAMETEQLGVEIFETADCEEN-IESQDRPKLEPFYVERYSWSQLKKLIADTRKYHGYM	59
gi	3513303	1	-----	1
70 80 90 100 110 120				
NOV64	49		VNKAPHDFQFVQKTDSEGPSHRLYYLCMPYGSRENSLLYSEIPKKVRKEALLLSWKQM	108
gi	17865311	49	VNKAPHDFQFVQKTDSEGPSHRLYYLCMPYGSRENSLLYSEIPKKVRKEALLLSWKQM	108
gi	17483229	1	-----	1
gi	12855335	61	MAKAPHDFMFVKRTDPDRPHSDRVVYLAMSGENRENTLFYSEIPKTIINRAAVLMLSWKPI	120
gi	11095188	60	MAKAPHDFMFVKRNDPDGPHSDRIYYLAMSGENRENTLFYSEIPKTIINRAAVLMLSWKPI	119
gi	3513303	1	-----	1
130 140 150 160 170 180				
NOV64	109		LDHFQATPHHGVYSREEELLRERKRLGVFGITSYDFHSESGFLFOASNSLFHCRDGGKN	168
gi	17865311	109	LDHFQATPHHGVYSREEELLRERKRLGVFGITSYDFHSESGFLFOASNSLFHCRDGGKN	168
gi	17483229	1	-----	1
gi	12855335	121	LDLFQATLDYCMYSREEELLRERKRLGTVGIAAYDYHPSGSGFLFOAGSGIYHVKDGGPH	180
gi	11095188	120	LDLFQATLDYCMYSREEELLRERKRLGTVGIAASYDYHPSGSGFLFOAGSGIYHVKDGGPQ	179
gi	3513303	1	-----	1
190 200 210 220 230 240				
NOV64	169		GFMVSPMKLEIKTQCSGPRMDPKICPADPAFFSFINNSDLWVANIEETGEERRLTFC HQG	228
gi	17865311	169	GFMVSPMKLEIKTQCSGPRMDPKICPADPAFFSFINNSDLWVANIEETGEERRLTFC HQG	228
gi	17483229	1	-----MKLEIKTQCSGPRMDPKICPADPAFFSFINNSDLWVANIEETGEERRLTFC HQG	54
gi	12855335	181	GFTQOPLRENLVETSCPNIRMDPKICPADEWDIAFTHSNDIWLNSNIVTREERRLTIVHNE	240
gi	11095188	180	GFTQOPLRENLVETSCPNIRMDPKICPADEWDIAFTHSNDIWLNSNIVTREERRLTIVHNE	239
gi	3513303	1	-----	1
250 260 270 280 290 300				
NOV64	229		LSNVLDPPKSAGVATFVIOEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVI	288
gi	17865311	229	LSNVLDPPKSAGVATFVIOEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVI	288
gi	17483229	55	LSNVLDPPKSAGVATFVIOEEFDRFTGYWWCPTASWEGSEGLKTLRILYEEVDESEVEVI	114
gi	12855335	241	LAMMEEDPRSAGVATFVIOEEFDRYSGYWWCPTAERTPS-GGKILRILYEEVDESEVEVI	299
gi	11095188	240	LAMMEEDARSAGVATFVIOEEFDRYSGYWWCPTAERTPS-GGKILRILYEEVDESEVEVI	298
gi	3513303	1	-----	1
310 320 330 340 350 360				
NOV64	289		HVPSPALEBRKTDSDRYRPTGSKNPKIALKLAEFQTDGQCKIVSTQEKELVQPFSSLFPK	348
gi	17865311	289	HVPSPALEBRKTDSDRYRPTGSKNPKIALKLAEFQTDGQCKIVSTQEKELVQPFSSLFPK	348
gi	17483229	115	HVPSPALEBRKTDSDRYRPTGSKNPKIALKLAEFQTDGQCKIVSTQEKELVQPFSSLFPK	174
gi	12855335	300	HVTSPMLETRRADSFYRYPKTGTANPKVTFKMSIIVDAAGGIIDVIDKELVQPFHILFEG	359
gi	11095188	299	HVTSPMLETRRADSFYRYPKTGTANPKVTFKMSIIMIDAGRIIDVIDKELVQPFHILFEG	358
gi	3513303	1	-----IVSTQEKELVQPFSSLFPK	19
370 380 390 400 410 420				
NOV64	349		VEYIARAGWTRDGKYAWAMFLDRPQQWLQVLVLLPPALFIPSTENEEQRLASARAVPRNVQ	408
gi	17865311	349	VEYIARAGWTRDGKYAWAMFLDRPQQWLQVLVLLPPALFIPSTENEEQRLASARAVPRNVQ	408
gi	17483229	175	VEYIARAGWTRDGKYAWAMFLDRPQQWLQVLVLLPPALFIPSTENEEQRLASARAVPRNVQ	234
gi	12855335	360	VEYIARAGWTPEGKHAWSILLDRSOTHLQIVLISPELFIPVEDDAMDORLIESVDPDSVT	419
gi	11095188	359	VEYIARAGWTPEGKYAWSILLDRSOTRLQIVLISPELFIPVEDDVMERORLIESVDPDSVT	418
gi	3513303	20	VEYIARAG-----AWAMFLDRPQQWLQVLVLLPPALFIPSTENEEQRLASARAVPRNVQ	72
430 440 450 460 470 480				
NOV64	409		PYVVYEEVTNVWINVHDIYFPFPQSEGEDELCFLRANECKTGFCFLYKVTAVLKSQGYDW	468
gi	17865311	409	PYVVYEEVTNVWINVHDIYFPFPQSEGEDELCFLRANECKTGFCFLYKVTAVLKSQGYDW	468
gi	17483229	235	PYVVYEEVTNVWINVHDIYFPFPQSEGEDELCFLRANECKTGFCFLYKVTAVLKSQGYDW	294
gi	12855335	420	PLTIYEETTDIWINIHDIHVFVFPQTH-EEDEEFIFASECKTGFRHLYKITISILKESKYKR	478
gi	11095188	419	PLTIYEETTDIWINIHDIHVFVFPQSH-EEDEEFIFASECKTGFRHLYKITISILKESKYKR	477
gi	3513303	73	PYVVYEEVTNVWINVHDIYFPFPQSEGEDELCFLRANECKTGFCFLYKVTAVLKSQGYDW	132
490 500 510 520 530 540				

NOV64	469	SEPFSPGDEDFKCPIKEEIALTSGEWEVLARHGSKIWNNEETKLVFYFGTKDTPLEHHLY	528
gi 17865311	469	SEPFSPGDEDFKCPIKEEIALTSGEWEVLARHGSKIWNNEETKLVFYFGTKDTPLEHHLY	528
gi 17483229	295	SEPFSPGDEDFKCPIKEEIALTSGEWEVLARHGSKIWNNEETKLVFYFGTKDTPLEHHLY	354
gi 12855335	479	SSGGLPAPSDFKCPIKEEIALTSGEWEVLGRHGSNIWVDEARLKVYFEGTKDSPLEHHLY	538
gi 11095188	478	SSGGLPAPSDFKCPIKEEIALTSGEWEVLGRHGSNIWVDEVRRLLVYFEGTKDSPLEHHLY	537
gi 3513303	133	SEPFSPGEG-----EQSLTN-----AIWNNEETKLVFYFGTKDTPLEHHLY	173
<div> <div>550560570580590600</div> <div>..... </div> </div>			
NOV64	529	VVSYEAAGEIVRLTTPGFHSHSCSMSONFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQ	588
gi 17865311	529	VVSYEAAGEIVRLTTPGFHSHSCSMSONFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQ	588
gi 17483229	355	VVSYEAAGEIVRLTTPGFHSHSCSMSONFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQ	414
gi 12855335	539	VTSYANPGEVRLTDRGYSHSCCLSRHCFDFFTSKYSNOKNPHCVSLYKLSPEDDDPVHKT	598
gi 11095188	538	VVSYVNPGEVRLTDRGYSHSCCLSRHCFDFFTSKYSNOKNPHCVSLYKLSPEDDDPVHKT	597
gi 3513303	174	VVSYEAAGEIVRLTTPGFHSHSCSMSONFDMFVSHYSSVSTPPCVHVYKLSGPDDDPLHKQ	233
<div> <div>610620630640650660</div> <div>..... </div> </div>			
NOV64	589	PRFWASMMEEAASCPDYPVPEIFHFHTRSDVRLYGMLYKPHALQPEKKHPTVLFVYGGPQ	648
gi 17865311	589	PRFWASMMEEAASCPDYPVPEIFHFHTRSDVRLYGMLYKPHALQPEKKHPTVLFVYGGPQ	648
gi 17483229	415	PRFWASMMEEAASCPDYPVPEIFHFHTRSDVRLYGMLYKPHALQPEKKHPTVLFVYGGPQ	474
gi 12855335	599	KEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYGGPQ	658
gi 11095188	598	KEFWATILDSAGPLPDYTPPEIFSFESTTGFTLYGMLYKPHDLQPGKKYPTVLFYGGPQ	657
gi 3513303	234	PRFWASMMEEA-----KIFHFHTRSDVRLYGMLYKPHALQPEKKHPTVLFVYGGPQ	284
<div> <div>670680690700710720</div> <div>..... </div> </div>			
NOV64	649	VQLVNNSFKGIKYLRNLNTLASLGYAVVVVIDGRGSCQRLRFEGALKNQMGQVEIEDQVEG	708
gi 17865311	649	VQLVNNSFKGIKYLRNLNTLASLGYAVVVVIDGRGSCQRLRFEGALKNQMGQVEIEDQVEG	708
gi 17483229	475	VQLVNNSFKGIKYLRNLNTLASLGYAVVVVIDGRGSCQRLRFEGALKNQMGQVEIEDQVEG	534
gi 12855335	659	VQLVNNRFKGVKYERLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQTEIDDQVEG	718
gi 11095188	658	VQLVNNRFKGVKYERLNTLASLGYVVVVVIDNRGSCHRGLKFEGAFKYKMGQTEIDDQVEG	717
gi 3513303	285	VQLVNNSFKGIKYLRNLNTLASLGYAVVVVIDGRGSCQRLRFEGALKNQMGQVEIEDQVEG	344
<div> <div>730740750760770780</div> <div>..... </div> </div>			
NOV64	709	LQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTE	768
gi 17865311	709	LQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTE	768
gi 17483229	535	LQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTE	594
gi 12855335	719	LQYLASRYDFIDLRVGIHGWSYGGYLSLMALMORSDFRVAIAGAPVTLMWHYDTGYTE	778
gi 11095188	718	LQYLASRYDFIDLRVGIHGWSYGGYLSLMALMORSDFRVAIAGAPVTLMWHYDTGYTE	777
gi 3513303	345	LQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHKPQVFKVAIAGAPVTVMAYDTGYTE	404
<div> <div>790800810820830840</div> <div>..... </div> </div>			
NOV64	769	RYMDVPENNOHGYEAGSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLVSQIRACK	828
gi 17865311	769	RYMDVPENNOHGYEAGSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLVSQIRACK	828
gi 17483229	595	RYMDVPENNOHGYEAGSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLVSQIRACK	654
gi 12855335	779	RYMGHPDQNEQGYLGSVAMQAEKEFPSEPNRLLILHGFLDENVHFAHTSILLSFLVRACK	838
gi 11095188	778	RYMGHPDQNEQGYLGSVAMQAEKEFPSEPNRLLILHGFLDENVHFAHTSILLSFLVRACK	837
gi 3513303	405	RYMDVPENNOHGYEAGSVALHVEKLPNEPNRLLILHGFLDENVHFFHTNFLVSQIRACK	464
<div> <div>850860870880890</div> <div>..... </div> </div>			
NOV64	829	PYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQEYL-----	863
gi 17865311	829	PYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQEYL-----	863
gi 17483229	655	PYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQEYL-----	689
gi 12855335	839	PYDL-----QIYPOERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI	883
gi 11095188	838	PYDL-----QIYPOERHSIRVPESGEHYELHLLHYLQENLGSRIAALKVI	882
gi 3513303	465	PYQLQVALPPVSPQIYPNERHSIRCPESGEHYEVTLLHFLQEYL-----	508

Table 64F lists the domain description from DOMAIN analysis results against NOV64. This indicates that the NOV64 sequence has properties similar to those of other proteins known to contain this domain.

Table 64F. Domain Analysis of NOV64					
gnl Pfam pfam00930, DPPIV_N_term, Dipeptidyl peptidase IV (DPP IV) N-terminal region. This family is an alignment of the region to the N-terminal side of the active site. The Prosite motif does not correspond to this Pfam entry.					
SEQ ID NO:863					
CD-Length = 504 residues					
Score = 112 bits (280), Expect = 9e-26					
NOV64:	200	FFSFINNSDLWVANIETGEERRLTFCHQGLSNVLDPPKSAGVATFVIQEEFDRFTG-YWW	258		
		+F+ +++L++ + +G ++T G SN + + G+ +V +EE WW			
Sbjct:	122	KLAFVRDNNLYIQKLPSGPAIQITT--DGKSNDIFN----GIPDWVYEEEEILSTDYALWW	175		
NOV64:	259	CPTASWEGSEGLKTLRLIYEEVDESEVEVIHVPSPALEE--RKTDsYRYPRTGSKNPKIA	316		
		P + Y ++SEV VI P + + +YP+ G NP +			
Sbjct:	176	SPDGD-----FLAYLRFNDSEVPVIEYPFYTDDSQYPEDMEIKYPKAGDPNPTVK	225		
NOV64:	317	LKLAEFQTDsQCKIVSTQEKELVQPFSSLFKVEYIARAGWTRDGKYAWAMFLDRPQQWL	376		
		L + G VS + +SL YI R W + + A +L+R Q			
Sbjct:	226	LFVVNLAD---GASVSE-----IPLPASLASGDYYITRVAWVTNERLA-VQWLNRDQNIS	276		
NOV64:	377	QLVLLPALFIPSTENEEQRLASARAVPRNVQPYVVYEEVTNVWINVHDIFYPFPQSEGE	436		
		L L A +S V +N +E+ W+ + P +G			
Sbjct:	277	VLSLCDTA-----SSTWNVVKN-----FEDSETGWVETFNPSLPVFPDGL	317		
NOV64:	437	DELCLFRANECKTGfCHLYKVTAVLKSQGYDWSEPFSPGEDEFKCPiKEEIALTSGEWEV	496		
		+L ++ + G+ HL E + K PI ALT G WEV			
Sbjct:	318	SY--YLDISD-RDGYKHLAYY-----ELDGKEPI----ALTkGNWEV	352		
NOV64:	497	LARHGSKIWNNEETKLvYFQGTkDTPLEHHLYVVSy---EAAGEIVRLTTPGFSHSCSM	552		
		+ + V+ +T VYF T++ E HLY +S + G+ +S S			
Sbjct:	353	T----NILGVDSKTDTVYFTATEEGSGERHLYSISLKGKkTTLSCQLDSERCGY-YsASF	407		
NOV64:	553	SQNFDMFVSHYSSVSTPPCVHVYKLsGPDdDPLHKQPRFWASMME-----AASCPDPYV	606		
		S N ++ YS P S D L +E A			
Sbjct:	408	SPNAKYIILTYSGPGVP---IQTLHSSNDTKELR-----TLEDNEALKKALKKNYQLP	456		
NOV64:	607	PPEIFHFHTRSDVRLYGMiYKPhALQPEKKHPTVLFVYGGPQVQLV	652		
		E + L + KP P KK+P + FVYGGP Q V			
Sbjct:	457	SKEFGKIKLADGITLNYQMIKPANFDPSKKYPVLFFVYGGPGSQQV	502		

NOV64 is a member of the family of dipeptidyl peptidases (DPPs). This group of enzymes catalyzes the removal of dipeptides from the N termini of polypeptides. This novel gene has greatest homology to a recently discovered protein, DPP8 (Abbott et al., Eur J Biochem 2000 Oct;267(20):6140-50). DPP8 in turn is related to DPP4, which is a cell surface peptidase involved in T-cell activation (Kahne et al., Int J Mol Med 1999 Jul;4(1):3-15). Other members of the peptidase family have been targeted as putative drug targets, for instance, in situations where they might cleave polypeptides beneficial in the prevention or reduction of a disease condition.

The NOV64 disclosed in this invention is predicted to be expressed in at least the following tissues: bone, bone marrow, brain (cerebellum, substantia nigra, thalamus), bronchus, cartilage, cervix, chorionic villus, coronary artery, colon, breast, heart, kidney, liver,

lung, lymph node, lymphoid tissue, ovary, placenta, pituitary gland, respiratory bronchiole, retina, skeletal muscle, skin, small intestine, spinal cord, spleen, testis, thymus, thyroid, umbilical vein, urinary bladder, vulva, adrenal gland/suprarenal gland, synovium/synovial membrane, and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV64 is provided in Example 2.

The nucleic acids and proteins of NOV64 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for treatment of patients suffering from: Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoarthritis, and other diseases, disorders and conditions of the like. A NOV64 nucleic acid encoding the dipeptidyl peptidase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dipeptidyl peptidase-like protein includes the nucleic acid whose sequence is provided in Table 64A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 64A while still encoding a protein that maintains its dipeptidyl peptidase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 64A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 37% of the bases may be so changed.

The novel protein of the invention includes the dipeptidyl peptidase-like protein whose sequence is provided in Table 64B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 64B while still encoding a protein that maintains its dipeptidyl peptidase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 40% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV65

NOV65 includes two dual specificity phosphatase-like proteins, designated herein as NOV65a and NOV65b.

NOV65a

A disclosed NOV65a (alternatively referred to herein as CG56651-01) includes the 711 nucleotide sequence (SEQ ID NO:) shown in Table 65A. A NOV65a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 652-654. The disclosed NOV65a maps to human chromosome 1.

Table 65A. NOV65a Nucleotide Sequence (SEQ ID NO:227)

```
ATGCTTCCCAAACGCGTGAGGGAGAAGATGGATGACACCAGCCTCTATAATACGCCCTGT
GTCCTGGACCTACAGCGGGCCCTGGTTTCAGGATCGCCAAGAGGCGCCCTGGAATGAGGTG
GATGAGGTCTGGCCCAATGTCTTCATAGCTGAGAAGAGTGTGGCTGTGAACAAGGGGAGG
CTGAAGAGGCTGGGAATCACCCACATTCTGAATGCTGCGCATGGCACCGGCGTTTACACT
GGCCCCGAATTCTACACTGGCCTGGAGATCCAGTACCTGGGTGTAGAGGTGGATGACTTT
CCTGAGGTGGACATTTCCAGCATTTCGGAAGGCGTCTGAGTTCCTGGATGAGGCGCTG
CTGACTTACAGAGGGAAAGTCCTGGTCAGCAGCGAAATGGGCATCAGCCGGTCAGCAGTG
CTGGTGGTCGCTACCTGATGATCTTCCACAACATGGCCATCCTGGAGGCTTTGATGACC
GTGCGTAAGAAGCGGGCCATCTACCCCAATGAGGGCTTCCTGAAGCAGCTGCGGGAGCTC
AATGAGAAGTTGATGAGGAGAGAGAAGAGGACTATGGCCGGGAGGGGGATCAGCTGAGG
CTGAGGAGGGCGAGGGCACTGGGAGCATGCTCGGGGCCAGAGTGCACGCCCTGACGGTGG
AAGAGGAGGACGACAGCGCCAGCCACCTGAGTGGCTCCTCCCTGGGGAAGG
```

The NOV65a polypeptide (SEQ ID NO:228) encoded by SEQ ID NO:227 is 217 amino acids in length and is presented using the one-letter amino acid code in Table 65B. The Psort profile for NOV65a predicts that this sequence is likely to be a Type II membrane protein, and is likely to be localized at the plasma membrane with a certainty of 0.4400. In

alternative embodiments, a NOV65a polypeptide is located to the endoplasmic reticulum (membrane) with a certainty of 0.8500, or to the nucleus with a certainty of 0.7400.

Table 65B. NOV65a Polypeptide Sequence (SEQ ID NO:228)

MLPKRVREKMDDTSLYNTPCVLDLQRALVQDRQEPWNEVDEVPNVFIAEKSVAVNKGR
LKRLGITHILNAAHGTGVYTGPEFYTGLEIQYLGEVDDFPEVDISQHFRKASEFLDEAL
LTYRGKVLVSSEMGISRSAVLVVAYLMIFHNMAILEALMTVRKKRAIYPNEGFLKQLREL
NEKLMRREKRTMAGRGDQLRLRRARALGACSGPECTP

5 *NOV65b*

The disclosed NOV65b (alternatively referred to herein as CG56652-02) includes the 3212 nucleotide sequence (SEQ ID NO:) shown in Table 65C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 3193-3195. The disclosed NOV65b maps to human chromosome 1.

10

Table 65C. NOV65b Nucleotide Sequence (SEQ ID NO:229)

ATGCTTCCCAAACGCGTGAGGGAGAAGATGGATGACACCAGCCTCTATAATACGCCCTGT
GTCTGGACCTACAGCGGGCCCTGGTTTCAGGATCGCCAAGAGGCGCCCTGGAATGAGGTG
GATGAGGTCTGGCCCAATGTCTTCATAGCTGAGAAGAGTGTGGCTGTGAACAAGGGAGG
CTGAAGAGGCTGGGAATCACCACATTTCTGAATGCTGCGCATGGCACCAGCGCTTTACACT
GGCCCCGAATTTCTACACTGGCCTGGAGATCCAGTACCTGGGTGTAGAGGTGGATGACTTT
CCTGAGGTGGACATTTCCAGCATTTCCGGAAGGCGTCTGAGTTCCTGGATGAGGCGCTG
CTGACTTACAGAGGAAAGTCTTGGTTCAGCAGCGAAATGGGCATCAGCCGGTCAGCAGTG
CTGGTGGTCCCTACCTGATGATCTTCCACAACATGGCCATCCTGGAGGCTTTGATGACC
GTGCGTAAGAAGCGGGCATCTACCCAATGAGGGCTTCTGAAGCAGCTGCGGGAGCTC
AATGAGAAGTTGATGGAGGAGAGAGAAGAGGACTATGGCCGGGAGGGGGATCAGCTGAG
GCTGAGGAGGGCGAGGGCACTGGGAGCATGCTCGGGGCCAGAGTGCACGCCCTGACGGTG
GAAGAGGAGGACGACAGCGCCAGCCACCTGAGTGGCTCCTCCCTGGGGAAGGCCACCCAG
GCCTCCAAGCCCTCACCTCATAGACGAGGAGGAGGAGGAGAACTGTACGAGCAGTGG
AAGAAGGGGCAGGGCTCCTCTCAGACAAGGTCCCCAGGATGGAGGTGGCTGGCGCTCA
GCCTCCTCTGGCCAGGGTGGGGAGGAGCTCGAGGACGAGGACGTGGAGAGGATCATCCAG
GAGTGGCAGAGCCGAAACGAGAGGTACCAAGCAGAAGGGTACCGAGGTGGGGAGGGAG
GAGGAGAAGGAGGAGGAGAGCGACGCTGGCTCCTCGGTGGGGAGGCGCGCGCACCCCTG
AGCGAGAGCAGCGCCTGGGAGAGCGTGAGCAGCCACGACATCTGGGTCTGAAGCAGCAG
CTGGAGCTGAACCGCCCGGACCACGGCAGGAGGCGCGCGCAGACTCGATGTCTCTGGAG
AGCACCTGGGACGCATGGAACGAGAGGCTGCTGGAGATTGAGAAGGAGGCTTCCCGGAGG
TACCACGCCAAGAGCAAGAGAGAGGAGGCGGCAGACAGGAGCTCAGAAGCAGGGAGCAGG
GTGCGGGAGGATGATGAGGACAGCGTGGGCTCTGAGGCCAGTTCTCTTCTACAACCTCTGC
AGCAGGAACAAGGACAAGCTCACTGCCCTGGAAGATGGAAGATCAAGAGAATCCAATTT
GGATTTCACAAGAAAGACTTGGGAGCGGGAGACAGCAGCGGTGAGCCCCGTGCAGAGGAG
GCAGTAGGGGAGAAGAACCCTCCGACGTGACCTGACAGCCTACCAGGCCCTGGAAGCTG
AAACACCAGAAGAAGGTGGGCAGTGAGAACAAGGAGGAGGTGGTGGAGCTCAGCAAGGGG
GAGGACTCGGCCCTTGGCTAAGAAGAGACAACGGAGGCTGGAGCTGCTGGAGAGAAGCCGG
CAGACGTGGAGGAGAGCCAGTCTATGGCAAGCTGGGAGGCGGACAGCTCCACGGCCAGC
GGGAGCATTCCCTGTCTGCGTTCTGGTCTGCAGACCCCTCAGTCAGCGCTGATGGGGAC
ACGACGTCACTAGTACAGCCAGAGCCACCGCTCCCACTGTCTCAGGCTGCAAGCAAC
ATAGCGGGGTGTTCAACCTCCAACCCACACACCCCTGCCTAACCTGCCAGTGGGGCCT
GGAGACACCATTTCCATTGCCAGTATCCAGAACTGGATTGCCAATGTAGTCAGTGAGACC
CTTGCTCAGAAGCAAAATGAAATGCTGCTGTTGTCCTGCTCACCGTCTGTGCAAGCATG
AAGCAGTACCAGCGGCTAGCTGCCCTGGGGGATGACCAAGTCTCCATGTCTAGTGGACAC
AGCAGCTCCTCCTTGGGTGGCTGCCTGTTGCCTCAGAGCCAGGCAAGACCCAGCTCTGAC
ATGCAGTCTGTGCTGTCTGCAACACCACACTGAGCTCACCCGGGAAAGTTGCAGAAGC
AAAGTGAGGGGGACAGCAAGCCCATCTTCAGCCTCTTGTGCTGACAATGTGGACCTAAAG
GAACTTGGCCGGAAGGAGAAGGAGATGCAGATGGAGCTTAGGGAGAAGATGTCTGAGTAC

```

AAAATGGAAGCTGGCCTCAGACAACAAACGCAGCTCCCTCTTCAAGAAGAAGAAGGTC
AAGGAAGATGAGGATGATGGTGTGGGTGATGGGGATGAGGACACTGACAGTGCCATAGGG
AGCTTCCGATATTCTTCCGCGAGTAATCCCAGAAACCTGAAACAGACACATGCTCCTCC
CTGGCTGTCTGTGATCACTATGCAAGTGGCAGCAGAGTTGGCAAAGAGATGGATAGCAGT
ATTAATAAGTGCTCAGTGGCCTCAGGACGGAGGAAAAACCTCCTTTCCAAAGTGACTGG
TCTGGAAGTTCCAGAGGGAAGTACACCAGATCGTCCCTGCTCAGGGAGACAGAGTCTAAA
TCCTCCAGTTACAAGTTTCCAAATCCCAGTCAGAGGAACAGGACACCTCCTCTACCAC
GAGGCAATGGCAACTCTGTAAGAAGCACTTCACGGTTCTCATCTTCTCCACCAGGGAG
GGCAGAGAGATGCACAAGTTCTCCAGGTCCACGTACAACGAGACCTCAAGTTCCCGAGAG
GAGAGCCCAGAGCCCTACTTCTTCCGCCGACCCAGAGTCTCAGAAAGGGAAGAGTCC
CCAGAACCACAGCGCCCAAATTGGGCCAGGTCCAGGGACTGGGAAGATGTGAAGAGTCA
TCCAAGTCAGACTTCTCTGAATTTGGAGCCAAGAGGAAGTTCACCCAGAGCTTTATGAGG
TCTGAAGAAGAGGGAGAGAAAGAGAGGACAGAAAACAGAGAAGAAGGGAGGTTTGCATCT
GGACGGCGGTCCCAGTATCGGAGAAGCACTGACAGGGAGGAAGAGGAAGAAATGGACGAT
GAAGCCATCATTGCTGCTTGGAGACGCCGCAAGAAGAAACAGGACCAAGCTGCAGAAA
AGGAGGGAGGACTGAGCTGGGGAAAATCTGAG
    
```

The NOV65b polypeptide (SEQ ID NO:230) encoded by SEQ ID NO:229 is 1064 amino acids in length and is presented using the one-letter amino acid code in Table 65D. The Psort profile for NOV65b predicts that this sequence is a Type II membrane protein, and is likely to be localized at the plasma membrane with a certainty of 0.7900. In alternative embodiments, a NOV65b polypeptide is located to Golgi bodies with a certainty of 0.3000 or to the nucleus with a certainty of 0.8200.

Table 65D. NOV65b Polypeptide Sequence (SEQ ID NO:230)

```

MLPKRVREKMDTSLYNTPCVLDLQRALVQDRQEAPWNEVDEVWPNVFIAEKSVAVNKRGR
LKRLGITHILNAAHGTGVYTGPEFYTGLEIQYLGVEVDDFPEVDISQHFRAKASEFLDEAL
LTYRGKVLVSSEMGISRSVAVLVVAYLMI FHNMAILEALMTVRKKRAIYPNEGFLKQLREL
NEKLMEEREEDYGREGGSAAEEGEGTGSMLGARVHALTVEEEDDSASHLSGSSLGKATQ
ASKPLTLIDEEEEKLYEQWKKGQGLLSDKVPQDGGGWSASSGQGGEELEDEDVERIIQ
EWQSRNERYQAEGYRRWGREEEKEEESDAGSSVGRRRRTLSESSAWESVSSHDIWVLKQQ
LELNRPDHGRRRRADSMSSSESTWDAWNERLLEIEKEASRRYHAKSKREEAADRSSEAGSR
VREDDSDVGSEASSFYNFCSRNDKLTALERWKIKRIQFGFHKDLGAGDSSGEPGAEE
AVGEKNPSDVSLTAYQAWLKHQKKVGSSENKEEVVELSKGEDSALAKKRQRRLLELLERSR
QTLEESQSMASWEADSSTASGSIPLSAFWSADPSVSADGDTTSLVSTQSHRSHLSQAASN
IAGCSTSNPTTLPNLPVPGDITISIASIQNWIANVSETLAQKQNEMLLLSRSPSVASM
KAVPAASCLGDDQVMSLSGHSSSSSLGGCLLPQSQARPSSDMQSVLSCNTTTLSPAESCRRS
KVRGTSKPIFSLFADNVLDLKEKGRKEKEMQELREKMSEYKMEKLASDNKRSLFKKKKV
KEDEDDGVGDGEDTDSAIGSFYSSRSNSQKPEPDTDCSSLAVCDHYASGSRVKGEMDSS
INKWLSGLRTEEKPPFQSDWSGSSRGKYTRSSLLRETESKSSSYKFSKSQSEEQDTSSYH
EANGNSVRSTSRFSSSSTREGREMHKFSRSTYNETSSSREESPEPYFFRRTPESSEREES
PEPQRPNWARSRDWEDVEESSKDFSEFGAKRKFTQSFMRSEEEGEKERTENREEGRFAS
GRRSQYRRSTDREEEEEEMDDEAIIAAWRRRQEETRTKLQKRRED
    
```

10

A BLAST analysis of NOV65 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV65 had high homology to other proteins as shown in Table 65E.

Table 65E. BLASTX results from PatP database for NOV65

			High Score	Smallest Sum Probability P (N)
Sequences producing High-scoring Segment Pairs:				
patp:AAE04836	Human SGP018 phosphatase polypeptide		858	1.5e-85
patp:AAB40919	Human ORFX ORF683 polypeptide sequence		767	6.5e-76
patp:AAE04837	Human SGP003 phosphatase polypeptide		410	4.4e-38
patp:AAV68779	Amino acid sequence of a human		389	7.4e-36
patp:AAB42334	Human ORFX ORF2098 polypeptide sequence		389	7.4e-36

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 328 of 531 bases (61%) identical to a gb:GENBANK-ID:AB027004|acc:AB027004.1 mRNA from *Homo sapiens* (mRNA for protein phosphatase).

- 5 The full amino acid sequence of the protein of the invention was found to have 80 of 174 amino acid residues (45%) identical to, and 115 of 174 amino acid residues (66%) similar to, the 198 amino acid residue ptmr:SPTREMBL-ACC:Q9UII6 protein from *Homo sapiens* (Human) (PROTEIN PHOSPHATASE). NOV65 also has homology to the other proteins shown in the BLASTP data in Table 65F.

10

Table 65F. NOV65 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14602535 gb AAH09778.1 AAH09778 (BC009778)	protein phosphatase [<i>Homo sapiens</i>]	198	81/178 (45)	117/178 (65)	3e-37
gi 17454087 ref XP_061191.1 (XM_061191)	similar to protein phosphatase (H. sapiens) [<i>Homo sapiens</i>]	370	82/186 (44)	121/186 (64)	4e-37
gi 7705959 ref NP_057448.1 (NM_016364)	protein phosphatase [<i>Homo sapiens</i>]	198	81/178 (45)	116/178 (64)	1e-36
gi 7305011 ref NP_038877.1 (NM_013849)	dual specificity phosphatase 13 [<i>Mus musculus</i>]	198	80/178 (44)	115/178 (63)	3e-36
gi 12839241 dbj BAB24480.1 (AK006247)	dual specificity phosphatase 13-putative [<i>Mus musculus</i>]	198	79/178 (44)	114/178 (63)	3e-35

This BLASTP data is displayed graphically in the ClustalW in Table 65G. A multiple sequence alignment is given, with the NOV65a and b proteins being shown on lines 1 and 2 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 65F.

15

Table 65G. ClustalW Alignment of NOV65

		790	800	810	820	830	840	
NOV65a	217						217
NOV65b	658	ASMKAVPAASCLGDDQVSMLSGHSSSSSLGGCLLPQSQARPSSDMQSVLSCNTTLSSPAES						717
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		850	860	870	880	890	900	
NOV65a	217						217
NOV65b	718	CRSKVRGTSKPIFSLFADNVDLKELGRKEKEMQMELREKMSEYKMEKLASDNKRSSLFKK						777
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		910	920	930	940	950	960	
NOV65a	217						217
NOV65b	778	KKVKEDEDDGVGDGDEDTDSAIGSFYSSRSNSQKPETDTCSSLAVCDHYASGSRVGKEM						837
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		970	980	990	1000	1010	1020	
NOV65a	217						217
NOV65b	838	DSSINKWLSGLRTEEKPPFQSDWSGSSRGKYTRSSLLRETESKSSSYKFSKQSEEQDTS						897
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		1030	1040	1050	1060	1070	1080	
NOV65a	217						217
NOV65b	898	SYHEANGNSVRSTSRFSSSSTREGREMHKFSRSTYNETSSSREESPEPYFFRRTPESSER						957
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		1090	1100	1110	1120	1130	1140	
NOV65a	217						217
NOV65b	958	EESPEPQRPNWARSRDWEDVEESSKSDFFSEFGAKRKFTQSFMRSEEEGEKERTENREEGR						1017
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 12839241	198	----- ----- ----- ----- ----- ----- ----- -----						198
		1150	1160	1170	1180			
NOV65a	217						217
NOV65b	1018	FASGRRSQYRRSTDREEEEMDDEAIIAAWRRRQEETRTKLQKRRED						1064
gi 14602535	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 17454087	370	----- ----- ----- ----- ----- ----- ----- -----						370
gi 7705959	198	----- ----- ----- ----- ----- ----- ----- -----						198
gi 7305011	198	----- ----- ----- ----- ----- ----- ----- -----						198

gi|12839241| 198 ----- 198

Table 65H lists the domain description from DOMAIN analysis results against NOV65. This indicates that the NOV65 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 65H. Domain Analysis of NOV65				
gnl Pfam pfam00782, DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region. SEQ ID NO:864				
CD-Length = 139 residues, 97.8% aligned				
Score = 113 bits (282), Expect = 1e-26				
NOV65:	42	EVWPNVFIAEKSVAVNKGRLKRLGITHILNAAHGTGVYTGPEFYTGLEIQYLGVEVDDFP	101	
		E+ P++++ A N L +LGITH++N F YL + VDD		
Sbjct:	4	EILPHLYLGSYPTASNLAFLSKLGITHVINVTVEVPNSKNSGF-----LYLHIPVDDNH	57	
NOV65:	102	EVDISQHFRAKASEFLDEALLTYRGKVLVSSEMGISRSASVLVVAYLMIFHNMAILEALMTV	161	
		E DIS + +A EF+++A GKVLV + GISRSA L++AYLM N+++ EA V		
Sbjct:	58	ETDISPYLDEAVEFIEDAR-QKGGKVLVHCQAGISRSATLIIAYLMKTRNLSLNEAYSFV	116	
NOV65:	162	RKKR-AIYPNEGFLKQLRELNEK	183	
		+++R I PN GF +QL E K		
Sbjct:	117	KERRPIISPNGFKRQLIEYERK	139	

The NOV65 gene of invention is a member of the family of dual specificity protein phosphatases (DSPs; Martell et al., Mol Cells 1998 Feb 28;8(1):2-11). DSPs recognize either serine/threonine (Ser/Thr) or tyrosine (Tyr) moieties as targets for dephosphorylation. These enzymes regulate mitogenic signal transduction and can thereby regulate the cell cycle. Some members of this family are effective tumor suppressors, for example, PTEN. PTEN is required during embryonic development and later in life, and mutations in this gene give rise to different kinds of inherited and sporadic cancers (Eng, Recent Prog Horm Res 1999;54:441-52; discussion 453). In Drosophila, members of the DSP family, such as puckered, have important roles in development (Martin-Blanco et al., Genes Dev 1998 Feb 15;12(4):557-70). The crystal structure of one member of the DSP family has been elucidated (Yuvaniyama et al., Science 1996 May 31;272(5266):1328-31) and this family has been successfully targeted for small molecule drug development (Ducruet et al., Bioorg Med Chem 2000 Jun;8(6):1451-66). In addition, overexpression of a DSP has been demonstrated to be a potential therapy for cardiac hypertrophy (Bueno et al., Circ Res 2001 Jan 19;88(1):88-96). NOV65 has closest homology to a phosphatase that is differentially regulated in the testis during spermatogenesis

and is therefore thought to be involved in sperm development and maturation (Nakamura et al., Biochem. J. 344 Pt 3, 819-825 (1999)).

The disclosed NOV65 is predicted to be expressed in at least the following tissues: heart, skeletal muscle, colon, fetal lung, head, and ovary. This information was derived by
5 determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV65 is provided in Example 2.

The NOV65 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example,
10 cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, fertility, polycystic ovarian syndrome, cancer, tissue degeneration, bacterial/viral/parasitic infection, systemic lupus erythematosus, autoimmune
15 disease, asthma, emphysema, scleroderma, allergy, ARDS, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, Hirschsprung's disease, Crohn's Disease, appendicitis as well as other diseases, disorders and conditions. The NOV65 nucleic acid encoding the phosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
20 assessed. The novel nucleic acid of the invention encoding a protein phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 65A or 65C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 65A or 65C while still encoding a protein that maintains its protein phosphatase-like activities and physiological functions, or a
25 fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence indicated in Table 65A or 65C, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes
30 nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense

binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 39% of the bases may be so changed.

The novel protein of the invention includes the protein phosphatase-like protein whose sequence is provided in Table 65B or 65D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 65B or 65D while still encoding a protein that maintains its protein phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 54% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV66

The disclosed NOV66 (alternatively referred to herein as CG56633-01) includes the 1036 nucleotide sequence (SEQ ID NO:231) shown in Table 66A. A NOV66 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 28-30 and ends with a stop codon at nucleotides 913-915. The disclosed NOV66 maps to human chromosome 3.

Table 66A. NOV66 Nucleotide Sequence (SEQ ID NO:231)

```
CCCGTCTCGCTGGGTTGTCCCGGTCCATGTATCTGGTCATGGTGCTGAGGAACCTGCTC
ATCATCCTGGCTGTCAGCTCTGACTCCCACCTCCACACCCCATGTAATCTTCTCTCTCC
AACCTGTGCTGGGCTGACATCGGTTTCACCTCGGCCATGGTTCCCAAGATGATTGTGGAC
ATGCAGTCTCATAGCAGAGTCATCTCTTATGCGGGCTGCTGACACGGATGTCTTTCTTG
GTCCTTTTTGTCATGTATAGAAGACATGCTCCTGACTGCGATGGCCTATGACTGCTTTGTA
GCCATCTGTGCGCCTCTGCACTACCCAGTCATCGTGAATCCTCACCTCTCTGTCTCTTA
GTTTTGGTGTCTTTTTCTTAGCCTGTTGGATTCCAGCTGCACAGTTTGATTGTGTTA
CAATTCACCTTCTCAAGAATGTGGAAATCTCTAATTTGTCTGTGAGCCATCTCAGCTT
CTCAACCTTGCCTGTTCTGACAGCGTCATCAATAGCATATTCTTATATTTGATAGTACT
ATGTTTGGTTTTCTTCCCATTTCAAGGATCCTTTTGTCTTACTATAAAATTGTCCCTCC
ATTCTAAGGATTTCATCGTCAGATGGGAAGTATAAAGCCTTCTCCACCTGTGGCTCTCAC
CTAGCAGTTGTTTGCTTATTTTATGGAACAGGCATTGGCGTGTACCTGACTTCAGCTGTG
TCACCACCCCCAGGAGTGGTGTGGTGGCGTCAGTGATGTACGCTGTGGTCACCCCCATG
CTGAACCCCTTTCATCTATAGCCTGAGAAACAGAGACATTCAAAGCGCCCTCTGGAGGCTG
CGCAGCAGAACAGTCAATCTCATGATCTGTTCCATCCTTTTTCTTGTGTGGGTAAGAAA
GGGCAACCACATTAAATCCCTGCATCTGCAAATCCTGCTCCTTAGTCACATTATTTTGTG
GGCTTGATGGCTTTTATTCCTTTCCGCATTTCTATGTGAATATTGTTTCTTCGTTATG
CCTTTAACTGGAATGG
```

A NOV66 polypeptide (SEQ ID NO:232) encoded by SEQ ID NO:231 is 295 amino acids in length and is presented using the one-letter amino acid code in Table 66B. The Psort profile for NOV66 predicts that this sequence is a Type IIIa membrane protein, has a signal

peptide and is likely to be localized at the plasma membrane with a certainty of 0.6400. In alternative embodiments, a NOV66 polypeptide is located to Golgi bodies with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.3700. The Signal P predicts a likely cleavage site for a NOV66 peptide is between positions 17 and 18, *i.e.*, at the dash in the sequence AVS-SD.

Table 66B. NOV66 Polypeptide Sequence (SEQ ID NO:232)

MYLVMVLRNLLIILAVSSDSLHTPMYFFLSNLCWADIGFTSAMVPKMIQSHSRVIS
YAGCLTRMSFLVLFACIEDMLLTAMAYDCFVAICRPLHYPVIVNPHLSVFLVLSFFLSL
LDSQLHSLIVLQFTFFKNVEISNFVCEPSQLNLACSDSVINSIFLYFDSTMFGFLPISR
ILLSYYKIVPSILRISSDGKYKAFSTCGSHLAVVCLFYGTGIGVYLTSAVSPPPRSGVV
ASVMYAVVTPMLNPFIIYSLNRDIQSALWRLRSRTVESHDLFHPFSCVGKKGQPH

A BLAST analysis of NOV66 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV66 had high homology to other proteins as shown in Table 66C.

Table 66C. BLASTX results from PatP database for NOV66

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P(N)
patp:AAG71875 Human olfactory receptor polypeptide	1466	5.6e-150
patp:AAE04583 Human G-protein coupled receptor-39	1377	1.5e-140
patp:AAU24551 Human olfactory receptor AOLFR38	1377	1.5e-140
patp:AAG71816 Human olfactory receptor polypeptide	1363	4.6e-139
patp:AAU24549 Human olfactory receptor AOLFR36	1354	4.1e-138

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 973 of 1036 bases (93%) identical to a gb:GENBANK-ID:AF042089|acc:AF042089.1 mRNA from *Homo sapiens* (chromosome 3, olfactory receptor pseudogene cluster 1, complete sequence, and myosin light chain kinase (MLCK) pseudogene, partial sequence). The full amino acid sequence of the protein of the invention was found to have 192 of 265 amino acid residues (72%) identical to, and 221 of 265 amino acid residues (83%) similar to, the 264 amino acid residue ptnr:SPTREMBL-ACC:O43789 protein from *Homo sapiens* (Human) (OLFACTORY RECEPTOR). NOV66 also has homology to the other proteins shown in the BLASTP data in Table 66D.

Table 66D. NOV66 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17466082 ref XP_070192.1 (XM_070192)	similar to olfactory receptor, family 7, subfamily C, member 3 (H. sapiens) [<i>Homo sapiens</i>]	349	221/250 (88)	229/250 (91)	e-111
gi 17482057 ref XP_064778.1 (XM_064778)	similar to G protein-coupled receptor homolog clone G3 (H. sapiens) [<i>Homo sapiens</i>]	251	220/250 (88)	229/250 (91)	e-105
gi 17448458 ref XP_070402.1 (XM_070402)	similar to OLFACTORY RECEPTOR 7C2 (OLFACTORY RECEPTOR 19-18) (OR19-18) (H. sapiens) [<i>Homo sapiens</i>]	528	221/242 (87)	221/242 (91)	e-101
gi 7443955 pir PC4369	olfactory receptor, HT2 - human (fragment)	264	192/265 (72)	221/265 (82)	6e-92
gi 4092819 gb AAD03353.1 (AC006271)	BC319430_5 [<i>Homo sapiens</i>]	263	191/264 (72)	220/264 (82)	44-91

This BLASTP data is displayed graphically in the ClustalW in Table 66E. A multiple sequence alignment is given, with the NOV66 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in

5 Table 66D.

Table 66E. ClustalW Alignment of NOV66	
NOV66	(SEQ ID NO:232)
gi 17466082	(SEQ ID NO:629)
gi 17482057	(SEQ ID NO:630)
gi 17448458	(SEQ ID NO:631)
gi 7443955	(SEQ ID NO:632)
gi 4092819	(SEQ ID NO:633)
<div> <div> 102030405060 </div> <div> </div> </div>	
NOV66	1
gi 17466082	1
gi 17482057	1
gi 17448458	1
gi 7443955	1
gi 4092819	1
<div> <div> 708090100110120 </div> <div> </div> </div>	
NOV66	1
gi 17466082	12
gi 17482057	1
gi 17448458	61
gi 7443955	1
gi 4092819	1
<div> <div> 130140150160170180 </div> </div>	

gi 4092819	232 YSLRNKDIQSALCRLHGRIIKSHHLHPFCYMG----- 263
------------	---

Table 66F lists the domain description from DOMAIN analysis results against NOV66. This indicates that the NOV66 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 66F. Domain Analysis of NOV66				
gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). SEQ ID NO:810				
CD-Length = 254 residues, 99.6% aligned Score = 88.6 bits (218), Expect = 5e-19				
NOV66:	9	NLLIILAVSSDSHLHTPMYFFLSNLCWADIGFTSAMVPKMIVDMQSHSRVISYAGCLTRM	68	
Sbjct:	2	NLL+IL + L TP FL NL AD+ F + P + + V A C	61	
NOV66:	69	SFLVLFACIEDMLLTAMAYDCFVAICRPLHYPIVNPVHLSVFLVLSVFFLSLLDSQLHSL	128	
Sbjct:	62	+ V+ +LLTA++ D ++AI PL Y I P + L+L+ + L+LL S L	121	
NOV66:	129	IVLQFTFFKNVEISNFCVCEPSQLNLACSDSVINSIFLYFDSTMFGFLPISRILLSYKYI	188	
Sbjct:	122	T + + P + + ++ + + LP+ IL+ Y +I	172	
NOV66:	189	VP-----SILRISSSDGKYKAFSTCGSHLAVVCLFYGTGIGVYL---TSAVSPPP	235	
Sbjct:	173	+ L+ SS + A + V + I + L ++	232	
NOV66:	236	RSGVVASVMYAVVTPMLNPFIY	257	
Sbjct:	233	+ ++ ++ A V LNP IY	254	

The olfactory system is able to distinguish several thousand odorant molecules. Olfactory receptors are believed to be encoded by an extremely large subfamily of G protein-coupled receptors. These receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors. They are responsible for the recognition and G protein-mediated transduction of odorant signals. The genes encoding these receptors are devoid of introns within their coding regions. Schurmans et al. (1993) cloned a member of this family of genes, OLFR1, from a genomic library by cross-hybridization with a gene fragment obtained by PCR. By isotopic in situ hybridization, they mapped the gene to 17p13-p12 with a peak at band 17p13. A minor peak was detected on chromosome 3, with a maximum in the region 3q13-q21. After MspI digestion, a RFLP was demonstrated. Using this in a study of CEPH pedigrees, they demonstrated linkage with D17S126 at 17pter-p12; maximum lod = 3.6 at theta = 0.0. Used as a probe on Southern blots under moderately stringent conditions, the cDNA hybridized to at least 3 closely related genes. Ben-Arie et al. (1994) cloned 16 human

OLFR genes, all from 17p13.3. The intronless coding regions are mapped to a 350-kb contiguous cluster, with an average intergenic separation of 15 kb. The OLFR genes in the cluster belong to 4 different gene subfamilies, displaying as much sequence variability as any randomly selected group of OLFRs. This suggested that the cluster may be one of several
5 copies of an ancestral OLFR gene repertoire whose existence may have predated the divergence of mammals. Localization to 17p13.3 was performed by fluorescence in situ hybridization as well as by somatic cell hybrid mapping.

The ability to distinguish different odors depends on a large number of different odorant receptors (ORs). Sullivan et al. (1996) noted that ORs are expressed by nasal olfactory
10 sensory neurons; each neuron expresses only 1 allele of a single OR gene. In the nose, different sets of ORs are expressed in distinct spatial zones. Neurons that express the same OR gene are located in the same zone; however, in that zone they are randomly interspersed with neurons expressing other ORs. This distribution suggested to the authors that, when the cell chooses an OR gene for expression, it may be restricted to a specific zonal gene set, but it may
15 select from that set by a stochastic mechanism. Proposed models of OR gene choice fall into 2 classes: locus-dependent and locus-independent. Locus-dependent models posit that OR genes are clustered in the genome, perhaps with members of different zonal gene sets clustered at distinct loci. In contrast, locus-independent models do not require that OR genes be clustered. To assess the feasibility of these models, Sullivan et al. (1996) determined the expression
20 zones, sequences, and chromosomal locations of a number of mouse OR genes. They mapped OR genes to 11 different regions on 7 chromosomes. These loci lie within paralogous chromosomal regions that appear to have arisen by duplications of large chromosomal domains followed by extensive gene duplication and divergence. These studies showed that OR genes expressed in the same zone map to numerous loci. Moreover, a single locus can
25 contain genes expressed in different zones. These findings raised the possibility that OR gene choice is locus-independent or involved consecutive stochastic choices.

Nekrasova et al. (1996) overexpressed human (OR17-4) and rat (olp4) olfactory receptor genes in insect cells, purified them, and characterized them biochemically. They identified monomeric, dimeric, and trimeric forms of the proteins corresponding to molecular
30 weights of 32, 69, and 94 kD by electrophoresis. The oligomers were resistant to reduction and alkylation and were therefore thought to be held together by SDS-resistant hydrophobic interactions, consistent with observations of other G protein-coupled receptors.

Glusman et al. (1996) described the results of complete sequencing of an OR-rich cosmid spanning the center of the OR gene cluster on 17p13.3. The resulting 40-kb sequence

revealed 3 known OR coding regions, 2 OR genes which may have originated from a tandem duplication event, and a new OR pseudogene fused to another OR gene.

Issel-Tarver and Rine (1996) characterized 4 members of the canine olfactory receptor gene family. The 4 subfamilies comprised genes expressed exclusively in olfactory epithelium.

5 Analysis of large DNA fragments using Southern blots of pulsed field gels indicated that subfamily members were clustered together, and that 2 of the subfamilies were closely linked in the dog genome. Analysis of the 4 olfactory receptor gene subfamilies in 26 breeds of dog provided evidence that the number of genes per subfamily was stable in spite of differential selection on the basis of olfactory acuity in scent hounds, sight hounds, and toy breeds.

10 Issel-Tarver and Rine (1997) performed a comparative study of 4 subfamilies of olfactory receptor genes first identified in the dog to assess changes in the gene family during mammalian evolution, and to begin linking the dog genetic map to that of humans. These 4 families were designated by them OLF1, OLF2, OLF3, and OLF4 in the canine genome. The subfamilies represented by these 4 genes range in size from 2 to 20 genes. They are all
15 expressed in canine olfactory epithelium but were not detectably expressed in canine lung, liver, ovary, spleen, testis, or tongue. The OLF1 and OLF2 subfamilies are tightly linked in the dog genome and also in the human genome. The smallest family is represented by the canine OLF1 gene. Using dog gene probes individually to hybridize to Southern blots of genomic DNA from 24 somatic cell hybrid lines. They showed that the human homologous OLF1
20 subfamily maps to human chromosome 11. The human gene with the strongest similarity to the canine OLF2 gene also mapped to chromosome 11. Both members of the human subfamily that hybridized to canine OLF3 were located on chromosome 7. It was difficult to determine to which chromosome or chromosomes the human genes that hybridized to the canine OLF4 probe mapped. This subfamily is large in mouse and hamster as well as human, so the rodent
25 background largely obscured the human cross-hybridizing bands. It was possible, however, to discern some human-specific bands in blots corresponding to human chromosome 19. They refined the mapping of the human OLF1 homolog by hybridization to YACs that map to 11q11. In dogs, the OLF1 and OLF2 subfamilies are within 45 kb of one another (Issel-Tarver and Rine (1996)). Issel-Tarver and Rine (1997) demonstrated that in the human OLF1 and
30 OLF2 homologs are likewise closely linked. By studying YACs, Issel-Tarver and Rine (1997) found that the human OLF3 homolog maps to 7q35. A chromosome 19-specific cosmid library was screened by hybridization with the canine OLF4 gene probe, and clones that hybridized strongly to the probe even at high stringency were localized to 19p13.1 and 19p13.2. These clones accounted, however, for a small fraction of the homologous human bands.

Rouquier et al. (1998) demonstrated that members of the olfactory receptor gene family are distributed on all but a few human chromosomes. Through fluorescence in situ hybridization analysis, they showed that OR sequences reside at more than 25 locations in the human genome. Their distribution was biased for terminal bands of chromosome arms. Flow-sorted chromosomes were used to isolate 87 OR sequences derived from 16 chromosomes. Their sequence relationships indicated the inter- and intrachromosomal duplications responsible for OR family expansion. Rouquier et al. (1998) determined that the human genome has accumulated a striking number of dysfunctional copies: 72% of these sequences were found to be pseudogenes. ORF-containing sequences predominate on chromosomes 7, 16, and 17.

Trask et al. (1998) characterized a subtelomeric DNA duplication that provided insight into the variability, complexity, and evolutionary history of that unusual region of the human genome, the telomere. Using a DNA segment cloned from chromosome 19, they demonstrated that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Three chromosomes appeared to have contained the sequence before humans migrated around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in chimpanzee and gorilla, that site being nonorthologous to any of the locations in the human genome. Three new members of the olfactory receptor (OR) gene family were found to be duplicated within this large segment of DNA, which was found to be present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. From its sequence, one of the OR genes in this duplicated block appeared to be potentially functional. The findings raised the possibility that functional diversity in the OR family is generated in part through duplications and interchromosomal rearrangements of the DNA near human telomeres.

Mombaerts (1999) reviewed the molecular biology of the odorant receptor (OR) genes in vertebrates. Buck and Axel (1991) discovered this large family of genes encoding putative odorant receptor genes. Zhao et al. (1998) provided functional proof that one OR gene encodes a receptor for odorants. The isolation of OR genes from the rat by Buck and Axel (1991) was based on 3 assumptions. First, ORs are likely G protein-coupled receptors, which characteristically are 7-transmembrane proteins. Second, ORs are likely members of a multigene family of considerable size, because an immense number of chemicals with vastly different structures can be detected and discriminated by the vertebrate olfactory system. Third, ORs are likely expressed selectively in olfactory sensory neurons. Ben-Arie et al. (1994) focused attention on a cluster of human OR genes on 17p, to which the first human OR

gene, OR1D2, had been mapped by Schurmans et al. (1993). According to Mombaerts (1999), the sequences of more than 150 human OR clones had been reported. The human OR genes differ markedly from their counterparts in other species by their high frequency of pseudogenes, except the testicular OR genes. Research showed that individual olfactory sensory neurons express a small subset of the OR repertoire. In rat and mouse, axons of neurons expressing the same OR converge onto defined glomeruli in the olfactory bulb.

Gilad et al. (2000) reported the population sequence diversity of genomic segments within a 450-kb cluster of olfactory receptor (OR) genes on chromosome 17. They found a dichotomy in the pattern of nucleotide diversity between OR pseudogenes and introns on the one hand and the closely interspersed intact genes on the other. They suggested that weak positive selection is responsible for the observed patterns of genetic variation. This was inferred from a lower ratio of polymorphism to divergence in genes compared with pseudogenes or introns, high nonsynonymous substitution rates in OR genes, and a small but significant overall reduction in variability in the entire OR gene cluster compared with other genomic regions. The dichotomy among functionally distinct segments within a short genomic distance requires high recombination rates within this OR cluster.

NOV66 is predicted to be expressed in at least the following tissues: lung, liver, ovary, spleen, testis. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV66 is provided in Example 2.

The nucleic acids and proteins of NOV66 are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, fertility, polycystic ovarian syndrome, cancer, tissue degeneration, bacterial/viral/parasitic infection, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis, Hirschsprung's disease, Crohn's Disease, appendicitis as well as other diseases, disorders and conditions. The NOV66 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A NOV66 nucleic acid of the invention encoding a Olfactory receptor-like protein includes the

nucleic acid whose sequence is provided in Table 66A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 66A while still encoding a protein that maintains its Olfactory receptor -like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence disclosed in Table 66A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 7% of the bases may be so changed.

The novel protein of the invention includes the olfactory receptor-like protein whose sequence is provided in Table 66B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 66B while still encoding a protein that maintains its Olfactory receptor -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 28% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV67

The disclosed NOV67 (alternatively referred to herein as CG56571-01) includes the 1072 nucleotide sequence (SEQ ID NO:233) shown in Table 67A. A NOV67 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 41-43 and ends with a stop codon at nucleotides 989-991. The disclosed NOV67 maps to human chromosome 7.

Table 67A. NOV67 Nucleotide Sequence (SEQ ID NO:233)

CATTCCCTCCTAACCACCTAGATTGAAGAAGTGAGGTTCAATGTCCCAACTGGGAAGGGA
CAACATAACCTGGGTGAGTGAGTTCACTCTAATGGGTCTCTCCAGTGACAGGCAGACCCA
GGCTGGACTCTTTTATCTTATTTTGGGGCTGCCTACCTGCTGACCCTGCTGGGCAATGGGCT
CATCTGCTCTTGATCTGCGCTGGACGTGAGACTCCACCTGCCCATGTATTCTTCTCTG
CAACCTCTCACTTGTGGACATCTGCTACACCTCCAGCAGGGTCCCTCAGATGCTGGTGCA
CTGCACCCCAAAAAGAACCATCTCTTTGGCCGATGTGGGACCCAGCTCTTTTCTC
CCTGGCCCTCGGAGGACCGAGCTTTTTGTGTGCTGGCCGAATGGCCTATGACCGTACGT
GGCTGTTTTCGACACCCCTGTGTATACATGACGTATGAGCCCAAGGCTCTGCATGGCACT
GGCAGCTGTCTCTTGCTAGTGGGCTGGCTAAATCTGCTATGGAGACGGCACTGACCAT
GCACCTGCCACCTGTGGGCAACAACGTGCTGAACCATGTGGCCTGTGAGACACTGGCACT
GGTCAGGTGGCGCTCTGGCATCACTTCTCAATCAGGTGGTCTATAGTGGCTCCAGTGT
GGTGGTGTCTGCTGGTGGCCTGCTGCCTGGTCTCGTGTCTTACACCTCATTGTAGTTGC
CGTCTCTGCAGATCCACTCCACCAGGGGCAACCGCAAGCCCTTTGACCTGTGGCTCCCA
CCTCACTGTGGTCTCCATATCTTATGGGATGGCCCTTTTACTCATGACGCTCAGCCTC
CATGGCCTGACCTGAGCAGGAAAAGGTGATGGTACTCTCTTATGCTGTGGTGACCCCAT
GTTGAATCTTTTATCTACAGTCTCGGGAACAAGGATGTGAAGGCAGCTCTGAGTCGAG
TCTGATGAGGAGCTCTGAATTAACCAATAGAGAGTGGTTTGTAGTAAACAAGAAGCCTCA
CTCTGAAAAACAGTGGGCATTGGCACTGTGCTCTCCAGTATAACGTGTGTACGC

A NOV67 polypeptide (SEQ ID NO:234) encoded by SEQ ID NO:233 is 312 amino acids in length and is presented using the one-letter amino acid code in Table 67B. The Psort profile for NOV67 predicts that this sequence is a Type IIIb membrane protein, has a signal peptide, and is likely to be localized at the plasma membrane with a certainty of 0.6000. In alternative embodiments, a NOV67 polypeptide is located to Golgi bodies with a certainty of 0.4000, to the endoplasmic reticulum (membrane) with a certainty of 0.3000, or to the mitochondrial membrane with a certainty of 0.3522. The Signal P predicts a likely cleavage site for a NOV67 peptide is between positions 58 and 59, *i.e.*, at the dash in the sequence VRL-HL.

Table 67B. NOV67 Polypeptide Sequence (SEQ ID NO:234)

MSQLGRDNITWVSEFILMGLSSDRQTQAGFLFLFGAAYLLTLLGNGLIILLIWLVDRLHL
PMYFFLCNLSVDICYTSSRRVQMLVHSKRTKIFSFARCGTQLFFSLALGGTFEFLLLA
MAYDRYVAVSDPLCYIIVMSPRLCMAHAIVSVLGLANSAMETALMHLPTCGHNVLNVH
ACETLALVRSACVDITFNQVVIVASSVVLLVPCCLVSLSYTLIVVAVLQIHSTQGHRKA
FGTCASHLTVVVISYGMALFTYMQPRSMASAEQEKMVLSYAVVTPMLNPFYISLRNKDV
KAALSRLMRSELKH

A BLAST analysis of NOV67 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV67 had high
15 homology to other proteins as shown in Table 67C.

Table 67C. BLASTX results from PatP database for NOV67

Smallest

1997年7月1日

564

Table 67F lists the domain description from DOMAIN analysis results against NOV67. This indicates that the NOV67 sequence has properties similar to those of other proteins known to contain this domain.

Table 67F. Domain Analysis of NOV67				
gnl Pfam pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). SEQ ID NO:810				
CD-Length = 254 residues, 93.3% aligned Score = 76.3 bits (186), Expect = 3e-15				
NOV67:	61	PMYFFLCNLSLVDICYTSSRVPQMLVHCTSKRKTISFARCGTQLFFSLALGGTEFLLLA	120	
Sbjct:	18	P FL NL++ D+ + + P L + A C + G LLL A	77	
NOV67:	121	MAYDRYVAVCDPLCYIAVMSPRLCMALAAVSWLVGLANSAMETALTMHLPTCGHNVLNHV	180	
Sbjct:	78	++ DRY+A+ PL Y + +PR L + W++ L S + N +	137	
NOV67:	181	ISIDRYLAIVHPLRYRIRTPRRAKVLILLVWVLALLSLPPLLFSWLRTVEEGNTTVCL	240	
Sbjct:	138	V+ + V ++ ++ V+++ L +L L+ S+ + A	197	
NOV67:	241	IDFPVESVKRSYVLLSTLVGFVLPPLLVLVCYTRILRTLKRKRSQSLKRRSSSERKAA	293	
Sbjct:	198	+ V + L + ++ L A V LNP IY	254	

5

The olfactory system is able to distinguish several thousand odorant molecules. Olfactory receptors are believed to be encoded by an extremely large subfamily of G protein-coupled receptors. These receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors. They are responsible for the recognition and G protein-mediated transduction of odorant signals. The genes encoding these receptors are devoid of introns within their coding regions. Schurmans et al. (1993) cloned a member of this family of genes, OLFR1, from a genomic library by cross-hybridization with a gene fragment obtained by PCR. By isotopic in situ hybridization, they mapped the gene to 17p13-p12 with a peak at band 17p13. A minor peak was detected on chromosome 3, with a maximum in the region 3q13-q21. After MspI digestion, a RFLP was demonstrated. Using this in a study of 3 CEPH pedigrees, they demonstrated linkage with D17S126 at 17pter-p12; maximum lod = 3.6 at theta = 0.0. Used as a probe on Southern blots under moderately stringent conditions, the cDNA hybridized to at least 3 closely related genes. Ben-Arie et al. (1994) cloned 16 human OLFR genes, all from 17p13.3. The intronless coding regions are mapped to a 350-kb contiguous cluster, with an average intergenic separation of 15 kb. The OLFR genes in the cluster belong to 4 different gene subfamilies, displaying as much sequence variability as any randomly selected group of OLFRs. This suggested that the cluster may be one of several

15

20

copies of an ancestral OLFR gene repertoire whose existence may have predated the divergence of mammals. Localization to 17p13.3 was performed by fluorescence in situ hybridization as well as by somatic cell hybrid mapping.

The ability to distinguish different odors depends on a large number of different odorant receptors (ORs). Sullivan et al. (1996) noted that ORs are expressed by nasal olfactory sensory neurons; each neuron expresses only 1 allele of a single OR gene. In the nose, different sets of ORs are expressed in distinct spatial zones. Neurons that express the same OR gene are located in the same zone; however, in that zone they are randomly interspersed with neurons expressing other ORs. This distribution suggested to the authors that, when the cell chooses an OR gene for expression, it may be restricted to a specific zonal gene set, but it may select from that set by a stochastic mechanism. Proposed models of OR gene choice fall into 2 classes: locus-dependent and locus-independent. Locus-dependent models posit that OR genes are clustered in the genome, perhaps with members of different zonal gene sets clustered at distinct loci. In contrast, locus-independent models do not require that OR genes be clustered. To assess the feasibility of these models, Sullivan et al. (1996) determined the expression zones, sequences, and chromosomal locations of a number of mouse OR genes. They mapped OR genes to 11 different regions on 7 chromosomes. These loci lie within paralogous chromosomal regions that appear to have arisen by duplications of large chromosomal domains followed by extensive gene duplication and divergence. These studies showed that OR genes expressed in the same zone map to numerous loci. Moreover, a single locus can contain genes expressed in different zones. These findings raised the possibility that OR gene choice is locus-independent or involved consecutive stochastic choices.

Nekrasova et al. (1996) overexpressed human (OR17-4) and rat (olp4) olfactory receptor genes in insect cells, purified them, and characterized them biochemically. They identified monomeric, dimeric, and trimeric forms of the proteins corresponding to molecular weights of 32, 69, and 94 kD by electrophoresis. The oligomers were resistant to reduction and alkylation and were therefore thought to be held together by SDS-resistant hydrophobic interactions, consistent with observations of other G protein-coupled receptors.

Glusman et al. (1996) described the results of complete sequencing of an OR-rich cosmid spanning the center of the OR gene cluster on 17p13.3. The resulting 40-kb sequence revealed 3 known OR coding regions, 2 OR genes which may have originated from a tandem duplication event, and a new OR pseudogene fused to another OR gene.

Issel-Tarver and Rine (1996) characterized 4 members of the canine olfactory receptor gene family. The 4 subfamilies comprised genes expressed exclusively in olfactory epithelium.

Analysis of large DNA fragments using Southern blots of pulsed field gels indicated that subfamily members were clustered together, and that 2 of the subfamilies were closely linked in the dog genome. Analysis of the 4 olfactory receptor gene subfamilies in 26 breeds of dog provided evidence that the number of genes per subfamily was stable in spite of differential selection on the basis of olfactory acuity in scent hounds, sight hounds, and toy breeds.

Issel-Tarver and Rine (1997) performed a comparative study of 4 subfamilies of olfactory receptor genes first identified in the dog to assess changes in the gene family during mammalian evolution, and to begin linking the dog genetic map to that of humans. These 4 families were designated by them OLF1, OLF2, OLF3, and OLF4 in the canine genome. The subfamilies represented by these 4 genes range in size from 2 to 20 genes. They are all expressed in canine olfactory epithelium but were not detectably expressed in canine lung, liver, ovary, spleen, testis, or tongue. The OLF1 and OLF2 subfamilies are tightly linked in the dog genome and also in the human genome. The smallest family is represented by the canine OLF1 gene. Using dog gene probes individually to hybridize to Southern blots of genomic DNA from 24 somatic cell hybrid lines. They showed that the human homologous OLF1 subfamily maps to human chromosome 11. The human gene with the strongest similarity to the canine OLF2 gene also mapped to chromosome 11. Both members of the human subfamily that hybridized to canine OLF3 were located on chromosome 7. It was difficult to determine to which chromosome or chromosomes the human genes that hybridized to the canine OLF4 probe mapped. This subfamily is large in mouse and hamster as well as human, so the rodent background largely obscured the human cross-hybridizing bands. It was possible, however, to discern some human-specific bands in blots corresponding to human chromosome 19. They refined the mapping of the human OLF1 homolog by hybridization to YACs that map to 11q11. In dogs, the OLF1 and OLF2 subfamilies are within 45 kb of one another (Issel-Tarver and Rine (1996)). Issel-Tarver and Rine (1997) demonstrated that in the human OLF1 and OLF2 homologs are likewise closely linked. By studying YACs, Issel-Tarver and Rine (1997) found that the human OLF3 homolog maps to 7q35. A chromosome 19-specific cosmid library was screened by hybridization with the canine OLF4 gene probe, and clones that hybridized strongly to the probe even at high stringency were localized to 19p13.1 and 19p13.2. These clones accounted, however, for a small fraction of the homologous human bands.

Rouquier et al. (1998) demonstrated that members of the olfactory receptor gene family are distributed on all but a few human chromosomes. Through fluorescence in situ hybridization analysis, they showed that OR sequences reside at more than 25 locations in the human genome. Their distribution was biased for terminal bands of chromosome arms. Flow-

sorted chromosomes were used to isolate 87 OR sequences derived from 16 chromosomes. Their sequence relationships indicated the inter- and intrachromosomal duplications responsible for OR family expansion. Rouquier et al. (1998) determined that the human genome has accumulated a striking number of dysfunctional copies: 72% of these sequences were found to be pseudogenes. ORF-containing sequences predominate on chromosomes 7, 16, and 17.

Trask et al. (1998) characterized a subtelomeric DNA duplication that provided insight into the variability, complexity, and evolutionary history of that unusual region of the human genome, the telomere. Using a DNA segment cloned from chromosome 19, they demonstrated that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Three chromosomes appeared to have contained the sequence before humans migrated around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in chimpanzee and gorilla, that site being nonorthologous to any of the locations in the human genome. Three new members of the olfactory receptor (OR) gene family were found to be duplicated within this large segment of DNA, which was found to be present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. From its sequence, one of the OR genes in this duplicated block appeared to be potentially functional. The findings raised the possibility that functional diversity in the OR family is generated in part through duplications and interchromosomal rearrangements of the DNA near human telomeres.

Mombaerts (1999) reviewed the molecular biology of the odorant receptor (OR) genes in vertebrates. Buck and Axel (1991) discovered this large family of genes encoding putative odorant receptor genes. Zhao et al. (1998) provided functional proof that one OR gene encodes a receptor for odorants. The isolation of OR genes from the rat by Buck and Axel (1991) was based on 3 assumptions. First, ORs are likely G protein-coupled receptors, which characteristically are 7-transmembrane proteins. Second, ORs are likely members of a multigene family of considerable size, because an immense number of chemicals with vastly different structures can be detected and discriminated by the vertebrate olfactory system. Third, ORs are likely expressed selectively in olfactory sensory neurons. Ben-Arie et al. (1994) focused attention on a cluster of human OR genes on 17p, to which the first human OR gene, OR1D2, had been mapped by Schurmans et al. (1993). According to Mombaerts (1999), the sequences of more than 150 human OR clones had been reported. The human OR genes differ markedly from their counterparts in other species by their high frequency of pseudogenes, except the testicular OR genes. Research showed that individual olfactory

sensory neurons express a small subset of the OR repertoire. In rat and mouse, axons of neurons expressing the same OR converge onto defined glomeruli in the olfactory bulb.

Gilad et al. (2000) reported the population sequence diversity of genomic segments within a 450-kb cluster of olfactory receptor (OR) genes on chromosome 17. They found a
5 dichotomy in the pattern of nucleotide diversity between OR pseudogenes and introns on the one hand and the closely interspersed intact genes on the other. They suggested that weak positive selection is responsible for the observed patterns of genetic variation. This was inferred from a lower ratio of polymorphism to divergence in genes compared with pseudogenes or introns, high nonsynonymous substitution rates in OR genes, and a small but
10 significant overall reduction in variability in the entire OR gene cluster compared with other genomic regions. The dichotomy among functionally distinct segments within a short genomic distance requires high recombination rates within this OR cluster.

NOV67 is predicted to be expressed in at least the following tissues: brain, testis, ovary, skeletal muscle, neuronal tissue. This information was derived by determining the
15 tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV67 is provided in Example 2.

The NOV67 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein. The
20 NOV67 nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding an olfactory receptor-like protein OLF3-like protein includes the nucleic acid whose sequence is provided in Table 67A, or a fragment thereof.

25 The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 67A while still encoding a protein that maintains its olfactory receptor-like protein OLF3-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence disclosed in Table 67A, including
30 nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the

TGCTCCAGTTAGACAGCTGCCACCTGGACTTGGGGCCTGAGGGAGGCACCGTGGAACCTCA
TCCAGGGCCGGGGCGGCCAAGGGCAACTGTGTGAGCCTGCTGTCCCAAGCCCCGAGGGTG
ACCCGCGCTTCAGCTTCCTCCTCCACTTCTACACAGTACCCATACCCAAAACCGGCACCC
TCAGCTGCACGGTAGCCCTGCGTCCCAAGACCGGGTCTCAAGACCAGGAAGTCCATAGGA
CTGTCTTCATGCGCTTGAACATCATCAGCCCTGACCTGTCTGGTTGCACAAGCAAAGGCC
TCGTCTGCCCCGCGTGCTGGGCATCACCTTTGGTGCCTTCCTCATCGGGGCCCTGTCTCA
CTGCTGCACTCTGGTACATCTACTCGCACACGCGTTCCCCCAGCAAGCGGGAGCCCCGTGG
TGGCGGTGGCTGCCCCGGCCTCCTCGGAGAGCAGCAGCACCAACCACAGCATCGGGAGCA
CCCAGAGCACCCCTGCTCCACAGCAGCATGGCATAGCCCCGGCCCCCGCGCTCGCCC
AGCAGGAGAGACTGAGCAGCCGCCAGCTGGGAGCACTGGTGTGAACTACCCCTGGGAGCC
AGTCTCCACTCGACCCAGAATGGAGCCTGCTCTCCGCGCCTACCCTTCCCGCCTCCCTC
TCAGAGGCCTGCTGCCAGTGCAGCCACTGGCTTGGAAACACCTTGGGGTCCCTCCACCCCA
CAGAACCTTCAACCCAGTGGGTCTGGGATATGGCTGCCCAGGAGACAGACCACTTGCCAC
GCTGTTGTAAAAACCAAGTCCCTGTTCATTGAACTGGATCCAGCACTGGTGAAGTGAAG
CTGGGCAGGAAGGGAGAAGTTGAAACAGATTGAGGCCAGCCAGCCAGGCCAACAGCACC
TCCCCGCTGGGAAGAGAAGAGGGCCAGCCAGAGCCACCTGGATCTATCCCTGCGGCCT
CCACACCTGAACTTGCCTAATACTGGCAGGGGAGACAGGAGCCTAGCGAGCCAGCC
TGGGAGCCCAGAGGGTGGCAAGAACAGTGGGCGTTGGGAGCCTAGCTCCTGCCACATGGA
GCCCCCTCTGCCGTCGGGCAGCCAGCAGAGGGGAGTAGCCAAGCTGCTTGTCTGGGC
CTGCCCTGTGTATTACCAACCAATAAATCAGACCATGAAACAGTGAAAAA

A NOV68 polypeptide (SEQ ID NO:236) encoded by SEQ ID NO:235 is 625 amino acids in length and is presented using the one-letter amino acid code in Table 68B. The Psort profile for NOV68 predicts that this sequence is a Type IIIa membrane protein, has a signal peptide, and is likely to be localized at the plasma membrane with a certainty of 0.6400. In alternative embodiments, a NOV68 polypeptide is located to Golgi bodies with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.3700. The Signal P predicts a likely cleavage site for a NOV68 peptide is between positions 25 and 26, *i.e.*, at the dash in the sequence SLA-ET.

Table 68B. NOV68 Polypeptide Sequence (SEQ ID NO:236)

MDRGTLPPLAVALLASCSLSPTSLAETVHCDLQPVGPERGEVTTYTTSQVSKGCVAPNA
ILEVHVLFLFPTGPSQLELTQASKQNGTWPREVLLVLSVNSSVFLHLQALGIPHLAY
NSSLVTFQEPGPVNTTELPSSSSSSLSFCMLEASQDMGRTLEWRPRTPALVRGCHLEGVA
GHKEAHLRLVLPGHSAGPRTVTVKVELSCAPGDLDAVLILQGPPYVSWLIDANHNMQIWT
TGEYSFKIFPEKNIRGFKLPDTPQGLLGEARMLNASIVASFVELPLASIVSLHASSCGGR
LQTSAPAIQTTPPKDTCSPPELLMSLIQTKCADDAMTLVLKELVAHLKCTITGLTFWDPS
CEAEDRGDKFVLRSAISSCGMQVSASMISNEAVVNILSSSSPQRKKVHCLNMDLSLFLG
LYLSPHFLQASNTIEPQQSFVQVRVSPSVSEFLQLDLSCHLDLGPPEGTVELIQGRAAK
GNCVSLSPSPGDPFRFSFLHFYTVPIPKTGTLSCTVALRPKTGSQDQEVHRTVFMRLN
IISPDLSGCTSKGLVLPVLGITFGAFLIGALLTAALWYIYSHTRSPSKREPVVAAPA
SSESSSTNHSIGSTQSTPCSTSSMA

A BLAST analysis of NOV68 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV68 had high homology to other proteins as shown in Table 68C.

Table 68C. BLASTX results from PatP database for NOV68

Smallest
Sum

Sequences producing High-scoring Segment Pairs:	High Score	Probability P(N)
patp:AAR54828 Endoglin - <i>Homo sapiens</i> , 658 aa.	2500	1.5e-259
patp:AAR99802 Endoglin - <i>Homo sapiens</i> , 658 aa.	2500	1.5e-259
patp:AAY82190 Human endoglin SEQ ID NO:2 - <i>Homo sapiens</i>	2500	1.5e-259
patp:AAR37808 Rat betaglycan - Synthetic, 853 aa.	215	2.2e-24
patp:AAR74601 Rat betaglycan contg. transforming growth factor	215	2.2e-24

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 2128 of 2128 bases (100%) identical to a gb:GENBANK-ID:HUMENDO|acc:J05481.1 mRNA from *Homo sapiens* (Human endoglin mRNA, 3' end).

- 5 The full amino acid sequence of the protein of the invention was found to have 509 of 624 amino acid residues (81%) identical to, and 531 of 624 amino acid residues (85%) similar to, the 658 amino acid residue ptnr:SWISSNEW-ACC:P17813 protein from *Homo sapiens* (Human) (ENDOGLIN PRECURSOR (CD105 ANTIGEN)). NOV68 also has homology to the other proteins shown in the BLASTP data in Table 68D.

10

Table 68D. NOV68 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 3041681 sp P17813 EGLN_HUMAN	ENDOGLIN PRECURSOR (CD105 ANTIGEN)	658	621/658 (94)	632/658 (94)	0.0
gi 15679936 gb AAH14271.1 AAH14271 (BC014271)	endoglin (Osler-Rendu-Weber syndrome 1) [<i>Homo sapiens</i>]	658	620/658 (94)	621/658 (94)	0.0
gi 105920 pir A36262	endoglin precursor - human	645	607/644 (94)	608/644 (94)	0.0
gi 4557555 ref NP_000109.1 (NM_000118)	endoglin precursor; Endoglin [<i>Homo sapiens</i>]	625	581/618 (94)	582/618 (94)	0.0
gi 6679649 ref NP_031958.1 (NM_007932)	endoglin [<i>Mus musculus</i>]	653	452/660 (68)	540/660 (75)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 68E. A multiple sequence alignment is given, with the NOV68 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in

- 15 Table 68D.

Table 68E. ClustalW Alignment of NOV68		
NOV68	(SEQ ID NO:236)	
gi 3041681	(SEQ ID NO:639)	
gi 15679936	(SEQ ID NO:640)	

gi	105920	(SEQ ID NO:641)	
gi	4557555	(SEQ ID NO:642)	
gi	6679649	(SEQ ID NO:643)	

			10	20	30	40	50	60	
NOV68	1	MDRGTLPLAVALLLASCSLSPTS-	LAETVHCDLQPVGPERGEV	TYTTSQVSKGCVAQAPN	59				
gi	3041681	1	MDRGTLPLAVALLLASCSLSPTS-	LAETVHCDLQPVGPERGEV	TYTTSQVSKGCVAQAPN	59			
gi	15679936	1	MDRGTLPLAVALLLASCSLSPTS-	LAETVHCDLQPVGPERGEV	TYTTSQVSKGCVAQAPN	59			
gi	105920	1	-----GASCSLSPTS-	LAETVHCDLQPVGPERGEV	TYTTSQVSKGCVAQAPN	46			
gi	4557555	1	MDRGTLPLAVALLLASCSLSPTS-	LAETVHCDLQPVGPERGEV	TYTTSQVSKGCVAQAPN	59			
gi	6679649	1	MDRGVLPLEPTLLFVIYSFVPTTGLAERV	CCDLQPVDPTRGEVTE	TTTSQVSEGCVAQAPN	60			

			70	80	90	100	110	120	
NOV68	60	AILEVHVLFLFPTGPSQLELT	LQASKQNGTWPREVLLVLSVNSSV	FHLHLQALGIPHLA	119				
gi	3041681	60	AILEVHVLFLFPTGPSQLELT	LQASKQNGTWPREVLLVLSVNSSV	FHLHLQALGIPHLA	119			
gi	15679936	60	AILEVHVLFLFPTGPSQLELT	LQASKQNGTWPREVLLVLSVNSSV	FHLHLQALGIPHLA	119			
gi	105920	47	AILEVHVLFLFPTGPSQLELT	LQASKQNGTWPREVLLVLSVNSSV	FHLHLQALGIPHLA	106			
gi	4557555	60	AILEVHVLFLFPTGPSQLELT	LQASKQNGTWPREVLLVLSVNSSV	FHLHLQALGIPHLA	119			
gi	6679649	61	AVREHVHVLFLDFPGMLSHLELT	LQASKQNGTETREVFLVLVS	NKNVVFVKFQAFIPLHLA	120			

			130	140	150	160	170	180	
NOV68	120	YNSSLVTFQEP	PGVNTTELPS-					SSSS	145
gi	3041681	120	YNSSLVTFQEP	PGVNTTELPSFP	KTQILEWAAERGPITSAAELND	PQSILLRLGQAQGS	178		
gi	15679936	120	YNSSLVTFQEP	PGVNTTELPSFP	KTQILEWAAERGPITSAAELND	PQSILLRLGQAQGS	178		
gi	105920	107	YNSSLVTFQEP	PGVNTTELPSFP	KTQILEWAAERGPITSAAELND	PQSILLRLGQAQGS	165		
gi	4557555	120	YNSSLVTFQEP	PGVNTTELPSFP	KTQILEWAAERGPITSAAELND	PQSILLRLGQAQGS	178		
gi	6679649	121	YDSSLVTFQGP	PRVNTIVLPSLTSRKQILD	WAATKGAITSIAALDD	PQSIVLQLGQDPKA	180		

			190	200	210	220	230	240	
NOV68	146	LSFCMLEASQDMGRTLEWR	PRTPALVRGCHLEGVAGHKEA	HILRVLPGHSAGPRTVT	TKV	205			
gi	3041681	179	LSFCMLEASQDMGRTLEWR	PRTPALVRGCHLEGVAGHKEA	HILRVLPGHSAGPRTVT	TKV	238		
gi	15679936	179	LSFCMLEASQDMGRTLEWR	PRTPALVRGCHLEGVAGHKEA	HILRVLPGHSAGPRTVT	TKV	238		
gi	105920	166	LSFCMLEASQDMGRTLEWR	PRTPALVRGCHLEGVAGHKEA	HILRVLPGHSAGPRTVT	TKV	225		
gi	4557555	179	LSFCMLEASQDMGRTLEWR	PRTPALVRGCHLEGVAGHKEA	HILRVLPGHSAGPRTVT	TKV	238		
gi	6679649	181	PFLCLPEAHKDMCATLEW	QSPRAQTPVQSCRLEGV	SGHKEAMILRLPGSEAG	PRTVTVM	240		

			250	260	270	280	290	300	
NOV68	206	ELSCAPGDLD	DAVLILQGP	PVSWLIDANHNMQI	WTTGEYSFKIFPEKNIR	GFKLPDTPQG	265		
gi	3041681	239	ELSCAPGDLD	DAVLILQGP	PVSWLIDANHNMQI	WTTGEYSFKIFPEKNIR	GFKLPDTPQG	298	
gi	15679936	239	ELSCAPGDLD	DAVLILQGP	PVSWLIDANHNMQI	WTTGEYSFKIFPEKNIR	GFKLPDTPQG	298	
gi	105920	226	ELSCAPGDLD	DAVLILQGP	PVSWLIDANHNMQI	WTTGEYSFKIFPEKNIR	GFKLPDTPQG	285	
gi	4557555	239	ELSCAPGDLD	DAVLILQGP	PVSWLIDANHNMQI	WTTGEYSFKIFPEKNIR	GFKLPDTPQG	298	
gi	6679649	241	ELSC	TSG--DAVLILHG	PPVSWFIDINHS	MTLTTGEYSVKIFPGSK	VKGVLPDTPQG	298	

			310	320	330	340	350	360	
NOV68	266	LLGEARMLNASIVAS	FVELPLASIVSLHASSCG	RLQTS	PAPIQTTPPKD	TCSP	ELLMSL	325	
gi	3041681	299	LLGEARMLNASIVAS	FVELPLASIVSLHASSCG	RLQTS	PAPIQTTPPKD	TCSP	ELLMSL	358
gi	15679936	299	LLGEARMLNASIVAS	FVELPLASIVSLHASSCG	RLQTS	PAPIQTTPPKD	TCSP	ELLMSL	358
gi	105920	286	LLGEARMLNASIVAS	FVELPLASIVSLHASSCG	RLQTS	PAPIQTTPPKD	TCSP	ELLMSL	345
gi	4557555	299	LLGEARMLNASIVAS	FVELPLASIVSLHASSCG	RLQTS	PAPIQTTPPKD	TCSP	ELLMSL	358
gi	6679649	299	LTAEARKLNASIV	TSFVELPLVSNVSL	RASSCGGV	FOTTPAPV	VTTPPKD	TCSP	ELLMSL

			370	380	390	400	410	420	
NOV68	326	IQTKCADDAM	TLVLKKELVAHLKCTIT	GLTFWD	PSCEADRGDKFV	LSAYSSCGM	QVSA	385	
gi	3041681	359	IQTKCADDAM	TLVLKKELVAHLKCTIT	GLTFWD	PSCEADRGDKFV	LSAYSSCGM	QVSA	418
gi	15679936	359	IQTKCADDAM	TLVLKKELVAHLKCTIT	GLTFWD	PSCEADRGDKFV	LSAYSSCGM	QVSA	418
gi	105920	346	IQTKCADDAM	TLVLKKELVAHLKCTIT	GLTFWD	PSCEADRGDKFV	LSAYSSCGM	QVSA	405
gi	4557555	359	IQTKCADDAM	TLVLKKELVAHLKCTIT	GLTFWD	PSCEADRGDKFV	LSAYSSCGM	QVSA	418
gi	6679649	359	IQPKCGNQVMT	LALNKKHVQTLQCTIT	GLTFWD	SSCAEDTD	DHLVLSSAYSSCGM	QVSA	418

			430	440	450	460	470	480	
--	--	--	-----	-----	-----	-----	-----	-----	--

NOV68:	414	--SLSFQLGLYLSPHF---	LQASNTIEPGQSFVQVRVSPSVSE-	FLQLDLSCHLDLGP-	466
		+ LY F Q+++T++ G + +		L + +C+ G	
Sbjct:	121	EGPPTCTYSLYKDDSF	GPSYQSADTVQLGDPVYHEW	SCDGRDDPSLGLLVHNCYATPGSD	180
NOV68:	467	--EGGTVELIQGRAAKGNCV	SLSPSPGDPFRFSFLH	FYTVPIPKTGTLSCTVALRPKT	524
		G +I + +SP F + +		C + L K+	
Sbjct:	181	PFSGPKYFIIDNGCPVD	RYLDSVSPYSSPSHYARF	SVKVFKFADRSLVYFHCQITL	CDKS 240
NOV68:	525	GSQD	528		
Sbjct:	241	DGSS	244		

Transforming growth factor-beta (TGF-beta) plays an important role in angiogenesis and vascular function. Endoglin, a transmembrane TGF-beta binding protein, is highly expressed on vascular endothelial cells and is the target gene for the hereditary haemorrhagic telangiectasia type I (HHT1), a dominantly inherited vascular disorder. The specific function of endoglin responsible for HHT1 is believed to involve alterations in TGF-beta responses. The initial interactions on the cell surface between endoglin and TGF-beta receptors may be an important mechanism by which endoglin modulates TGF-beta signalling, and thereby responses. On human microvascular endothelial cells, endoglin is co-expressed and is associated with betaglycan, a TGF-beta accessory receptor with which endoglin shares limited amino acid homology. This complex formation may occur in either a ligand-dependent or a ligand-independent manner. In addition, three higher order complexes containing endoglin, type II and/or type I TGF-beta receptors, also can occur on these cells. Thus endoglin may modify TGF-beta signalling by interacting with both betaglycan and the TGF-beta signalling receptors at physiological receptor concentrations and ratios (Wong et al., 2000, Eur J Biochem vol. 267 : 5550-60).

Endoglin is a homodimeric membrane glycoprotein. In association with transforming growth factor (TGF)-ss receptors I and II, endoglin can also bind TGF-ss1 and -ss3 and form a functional receptor complex. In human vascular tissue, endoglin immunolabeling is shown to be higher in endarterectomy specimens removed from diseased coronary arteries than in normal internal mammary arteries. In vitro, antisense oligonucleotides to endoglin is shown to decrease its expression and antagonized the TGF-ss-mediated inhibition of human and porcine SMC migration. Thus, upregulation of endoglin occurs during arterial repair and in established atherosclerotic plaques and may be required for modulation of SMC migration by TGF-ss (Ma X et al., 2000, Arterioscler Thromb Vasc Biol vol. 20 :2546-52).

Hereditary hemorrhagic telangiectasia (HHT) is an inherited autosomal dominant vascular dysplasia caused by mutations in either endoglin (HHT1) or activin-like kinase receptor-1 (ALK-1) (HHT2). The majority of the mutations in endoglin cause frameshifts and

premature stop codons. Although initial reports suggested a dominant-negative model for HHT1, more recent reports have suggested that mutations in endoglin lead to haploinsufficiency. Expression of the missense mutants alone revealed that they are misfolded and that most show no cell surface expression. When co-expressed with wild-type endoglin, the missense mutants are able to dimerize with the normal endoglin protein and are trafficked to the cell surface. Thus either dominant-negative protein interactions or haploinsufficiency can cause HHT1 (Lux et al., 2000, Hum Mol Genet vol 9 : 745-55).

NOV68 is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV68 is provided in Example 2.

The NOV68 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: arterial injuries, cerebral arteriovenous malformalities, pregnancy complications, carcinomas such as breast and mammary carcinoma as well as other diseases, disorders and conditions. The NOV68 nucleic acid encoding the endoglin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a Endoglin (CD105 antigen)-like protein includes the nucleic acid whose sequence is provided in Table 68A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 68A while still encoding a protein that maintains its Endoglin (CD105 antigen)-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence disclosed in Table 68A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications

include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

5 In the mutant or variant nucleic acids, and their complements, up to about 0% of the bases may be so changed.

The novel protein of the invention includes the Endoglin (CD105 antigen)-like protein whose sequence is provided in Table 68B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table
10 68B while still encoding a protein that maintains its Endoglin (CD105 antigen)-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 19% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic
15 methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV69

NOV69 includes two IL1-like proteins, designated herein as NOV69a and NOV69b.
20 These are splice variants of sequence accession number CG56950-01.

NOV69a

The disclosed NOV69a (alternatively referred to herein as CG56950-01) includes the 414 nucleotide sequence (SEQ ID NO:) shown in Table 69A. A NOV69a ORF begins with a
25 Kozak consensus ATG initiation codon at nucleotides 100-102 and ends with a TGA codon at nucleotides 412-414. The disclosed NOV69a maps to human chromosome 7.

**Table 69A. NOV69a Nucleotide Sequence (SEQ ID
NO:237)**

ATGGAAAAGCATTGAAAATTGACACACCTCAGCAGGGGAGCATTTCAGGATATCAATCAT CGGGTGTGGGTTCTTCAGGACCAGACGCTCATAGCAGTCCCGAGGAAGGACCGTATGTCT CCAGTCACTATTGCCTTAATCTCATGCCGACATGTGGAGACCCTTGAGAAAGACAGAGGG AACCCACACTGCAGCTGAAGGAAAAGGATATAATGGATTGTACAACCAACCCGAGCCT GTGAAGTCCTTTCTCTTCTACACAGCCAGAGTGGCAGGAATCCACCTTCGAGTCTGTG GCTTTCCTGGCTGGTTTCATCGCTGTGAGGAGGCTGCTCCTTCATCCTTACC CAAGAACTGGGGAAGCCAACACTACTGACTTTGGGTTAACTATGCTGTTTAA
--

A NOV69a polypeptide (SEQ ID NO:238) encoded by SEQ ID NO:237 is 137 amino acids in length and is presented using the one-letter amino acid code in Table 69B. The Psort profile for NOV69a predicts that this sequence has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV69a polypeptide is located to lysosomes with a certainty of 0.1514, or to peroxisomal microbodies with a certainty of 0.2384.

Table 69B. NOV69a Polypeptide Sequence (SEQ ID NO:238)

MEKALKIDTPQOGSIQDINHRVWVLQDQTLIAVPRKDRMSPVTIALISCRHVETLEKDRG
NPTLQLKEKDIMDLYNQPEPVKSFLFYHSQSGRNSTFESVAFPGWFIAVSSEGGCPLILT
QELGKANTTDFGLTMLF

NOV69b

The disclosed NOV69b (alternatively referred to herein as CG56136-02) includes the 411 nucleotide sequence (SEQ ID NO:) shown in Table 69C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 409-411. The disclosed NOV69b maps to human chromosome 2q12-14.1.

Table 69C. NOV69b Nucleotide Sequence (SEQ ID NO:239)

ATGGAAAAAGCATTGAAAGTTGACACACCTCAGCGGGGAGCATTGAGGATATCAATCAT
CGGGTGTGGGTCTTCAGGACCAGACGCTCATAGCAGTCCCGAGGAAGGACCGTATGTCT
CCAGTCACTATTGCCTTAATCTCATGCCGACATGTGGAGACCCCTTGAGAAAGACAGAGG
AACCCCATCTACCTGGGCCTGAATGGACTCAATCTCTGCCTGATGTGTGTTCAAGTCGGG
GACCAGCCCACTGCAGATGAACCAGAGTGGCAGGAACCTCCACCTTCGAGTCTGTGGCT
TTCCCTGGCTGGTTGATCGCTGTGAGTCTGAAGGAGGCTGTCTCTCATCCTTACCCAA
GAAGTGGGGAAGCCAACACTACTGACTTTGGGTAACTATGCTGTTTAA

A NOV69b polypeptide (SEQ ID NO:240) encoded by SEQ ID NO:239 is 136 amino acids in length and is presented using the one-letter amino acid code in Table 69D. The Psort profile for NOV69b predicts that this sequence has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500. In alternative embodiments, a NOV69b polypeptide is located to lysosomes with a certainty of 0.2305.

Table 69D. NOV69b Polypeptide Sequence (SEQ ID NO:240)

MEKALKVDTPQRGSIQDINHRVWVLQDQTLIAVPRKDRMSPVTIALISCRHVETLEKDRG
NPIYLGGLNGLNLCMLCVQVGDQPTLQMNQSGRNSTFESVAFPGWLIASSEGGCPLILTQ
ELGKANTTDFGLTMLF

A BLAST analysis of NOV69 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV69 had high homology to other proteins as shown in Table 69E.

5

Table 69E. BLASTX results from PatP database for NOV69			
		High Score	Smallest Sum Probability P(N)
Sequences producing High-scoring Segment Pairs:			
patp:AAW86286	Rodent interleukin (IL)-1 epsilon polypeptide	263	1.1e-38
patp:AAV24049	Amino acid sequence of a murine SPOIL-II	263	1.1e-38
patp:AAE06662	Mouse interleukin-1epsilon (IL-1epsilon)	263	1.1e-38
patp:AAV70217	Human Interleukin-1 epsilon protein	414	1.7e-38
patp:AAV70218	Human Interleukin-1 epsilon	414	1.7e-38

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 258 of 263 bases (98%) identical to a gb:GENBANK-ID:AF201831|acc:AF201831.1 mRNA from *Homo sapiens* (FIL1 epsilon mRNA). The full amino acid sequence of the protein of the invention was found to have 82 of 89 amino acid residues (92%) identical to, and 87 of 89 amino acid residues (97%) similar to, the 158 amino acid residue ptnr:SPTREMBL-ACC:Q9UHA7 protein from *Homo sapiens* (Human) (FIL1 EPSILON). NOV69 also has homology to the other proteins shown in the BLASTP data in Table 69F.

15

Table 69F. NOV69 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7657092 ref NP_055255.1 (NM_014440)	Interleukin-1 Superfamily 1 [<i>Homo sapiens</i>]	158	137/158 (86)	137/158 (86)	7e-73
gi 9506601 ref NP_062323.1 (NM_019450)	interleukin 1 family, member 6 (epsilon) [<i>Mus musculus</i>]	160	77/156 (49)	104/105 (66)	4e-38
gi 9665234 ref NP_062564.1 (NM_019618)	interleukin-1 homolog 1 [<i>Homo sapiens</i>]	169	76/147 (51)	97/147 (65)	2e-35
gi 7657090 ref NP_055253.1 (NM_014438)	Interleukin-1 Superfamily □ [<i>Homo sapiens</i>]	157	62/145 (42)	85/145 (57)	2e-27
gi 12844800 dbj BAB26505.1 (AK009787)	homolog to FIL1 ETA-putative [<i>Mus musculus</i>]	183	58/144 (40)	77/144 (53)	4e-23

This BLASTP data is displayed graphically in the ClustalW in Table 69G. A multiple sequence alignment is given, with the NOV69 protein being shown on line 1 in a ClustalW

$$-\frac{1}{\Gamma_0} \left(\frac{\partial \Gamma_0}{\partial t} + \nabla_{\parallel} \Gamma_0 v_{\parallel} \right) = -\frac{1}{\Gamma_0} \left(\frac{\partial \Gamma_0}{\partial t} + \nabla_{\parallel} \Gamma_0 v_{\parallel} \right)$$
[illegible]

580

Table 69H. Domain Analysis of NOV69

gnl Pfam pfam00340, IL1, Interleukin-1 / 18. This family includes interleukin-1 and interleukin-18. SEQ ID NO:1392				
CD-Length = 142 residues				
Score = 60.1 bits (144), Expect = 8e-11				
NOV69:	57	KDRGNPTLQLKEKDIMDLYNQPEPVKSFLFYHSQSGRNSTFESVAFPGWFIASSEGSCP	116	
		K+ P LQL+ + E K F F ++ G FES A+P WFIA E P		
Sbjct:	62	KEGDEPVLQLEMVEPPKYIKNSEMDKRFFFEKTEIGSKVYFESAAYPNWFIATKQEEDRP	121	
NOV69:	117	LILTQELGKANTTDF	131	
		+ L +++ TDF		
Sbjct:	122	VFLANGPPESDITDF	136	

There are two structurally distinct forms of IL1: IL1(alpha), which is the acidic form with pI5, and IL1(beta) (IL1B; 147720), the neutral form with pI7. Both are 17-kD proteins coded by separate genes. The IL1A gene has 10,206 bp with 7 exons and 6 introns (Furutani et al., 1986). By Southern transfer analysis of DNAs from human-rodent somatic cell hybrids, Modi et al. (1988) assigned the IL1A gene to chromosome 2. Regional localization to 2q13-q21 was achieved by in situ hybridization. Lafage et al. (1989) confirmed assignment to 2q13 by in situ hybridization.

The IL1A and IL1B proteins, which are synthesized by a variety of cell types including activated macrophages, keratinocytes, stimulated B lymphocytes, and fibroblasts, are potent mediators of inflammation and immunity. Lord et al. (1991) demonstrated that both the alpha and beta forms, but particularly the beta form, are transcribed in polymorphonuclear leukocytes stimulated with LPS. Both IL1A and IL1B stimulate osteoclast activity in vitro and are potent bone resorbing factors. Sabatino et al. (1988) studied the effects of 72-hour subcutaneous infusions of interleukins 1-alpha and -beta on plasma, calcium, and bone morphology. Both interleukins 1 caused a marked, dose-dependent increase in plasma calcium. Increased numbers of osteoclasts and bone resorption surfaces were observed on quantitative histomorphometry of bone. The results suggest a role for IL1 in the modulation of extracellular fluid calcium homeostasis. Hogquist et al. (1991) demonstrated that interleukin-1 is involved in apoptosis (cell death). Both the alpha and the beta forms are released as a consequence of cell injury regardless of the insult.

Bailly et al. (1993) elucidated a polymorphism that consists of a variable number of repeats of a 46-bp sequence within intron 6 of the IL1A gene. Among 72 unrelated persons, they identified 6 different alleles ranging from 5 to 18 repeats; the most frequent allele, present in 62%, contained 9 repeats. They suggested that the polymorphism may be of significance in gene function, since each repeat contains 3 potential binding sites for transcription factors.

Gray et al. (1986) showed that in the mouse also there are at least 2 interleukin-1 genes, IL1(alpha) and IL1(beta). Boulton et al. (1989) used in situ chromosome hybridization to show that the 2 IL1 genes in the mouse are located in the F region of chromosome 2. It had previously been shown by studies in mouse-hamster somatic cell hybrids and in recombinant inbred strains that the 2 genes are tightly linked on murine chromosome 2, approximately 4.7 cM distal to beta-2-microglobulin. By pulsed field gel electrophoresis, Silver et al. (1990) showed that the mouse IL1a and IL1b genes are contained in a genomic fragment of about 70 kb. Further studies suggested that IL1b lies 5-prime to IL1a, that the 2 genes are oriented in the same direction, and that they are separated by about 50 kb. From restriction mapping of the human genomic region, Nicklin et al. (1994) concluded that, relative to one terminal CpG island, the 3 genes mapped to the following intervals: IL1A was between +0 and +35 kb, IL1B between +70 and +110 kb, and IL1RN (147679) between +330 and +430 kb. Since the assignment of IL1RN to 2q14.2 appears to be the most definitive localization, the IL1A and IL1B genes can be presumably be said to be also on 2q14. Cox et al. (1998) carried out studies with multiallelic markers that grouped the 3 genes into a biallelic system for use in association studies. They identified a common, 8-locus haplotype of the IL1 gene cluster.

Hurwitz et al. (1992) studied the role of IL1 in the ovary, using a solution hybridization/RNase protection assay to test for expression of the IL1 gene, its type I receptor (IL1R; 147810), and its receptor antagonist (IL1RN). They presented findings which, taken together, revealed the existence of a complete, highly compartmentalized, hormone-dependent intraovarian IL1 system.

Since IL1 is an important cytokine in the control of the inflammatory response central to the pathology of rheumatoid arthritis (180300), Cox et al. (1999) used the combined sib-TDT (transmission/disequilibrium test) and TDT, in addition to parametric and nonparametric linkage methods, to investigate candidate genes of the IL1 gene cluster in the 2q13 region. Several tightly linked IL1 cluster markers yielded suggestive evidence for linkage in the combined TDT in those families in which affected sibs did not share 2 HLA-DRB1 alleles identical by descent. The evidence was significant in those with severe disease, as assessed by the presence of bone erosions. In contrast, there was no evidence of linkage using nonparametric linkage analysis, but parametric analysis revealed weak evidence of linkage when marker-trait disequilibrium was incorporated into the analysis. The data provided preliminary evidence for linkage of genes of the IL1 cluster to rheumatoid arthritis and suggested a possible role for this region in severe erosive disease.

Kornman et al. (1997) suggested that genetic polymorphisms of the IL1A and IL1B genes may be associated with severity of periodontitis in adult nonsmokers. The IL1B polymorphism was referred to as IL1B+3953 and the IL1A polymorphism was referred to as IL1A-889. Nonsmokers aged 40 to 60 carrying the '2' allele (in either homozygous or heterozygous state) at both loci were observed to have nearly 19 times the risk of developing severe periodontitis compared to subjects homozygous for the '1' allele at either or both of these loci. Because of the implication of interleukin-1 in adult periodontitis, Diehl et al. (1999) undertook an evaluation of the role of these IL1A and IL1B polymorphisms in early-onset periodontitis (EOP; see 170650) in 28 African-American families and 7 Caucasian-American families with 2 or more affected members. The 2 major EOP subtypes, localized juvenile periodontitis and generalized early-onset periodontitis, encompassing rapidly progressive periodontitis and generalized juvenile periodontitis, were analyzed separately and together. They obtained highly significant evidence of linkage disequilibrium for both groups of generalized EOP subjects. A similar trend was noted for the localized form. The IL1 alleles associated with high risk of EOP had been suggested previously to be correlated with low risk for severe adult periodontitis. Linkage disequilibrium with generalized EOP was equally strong for smoking and nonsmoking subjects. IL1A and IL1B polymorphisms were in strong linkage disequilibrium with each other in Caucasians but not in African Americans. Haplotype analyses evaluating both polymorphisms simultaneously indicated that the IL1B variant is likely to be more important for EOP risk. Sib pair linkage analyses, by contrast, provided only marginal support for a gene of very major effect on EOP risk attributable to these IL1 polymorphisms. Diehl et al. (1999) interpreted their results as indicating that EOP is a complex, oligogenic disorder, with interleukin-1 genetic variation contributing an important but not exclusive influence on disease risk.

NOV69 is predicted to be expressed in at least the following tissues: spleen, lymph node, thymus, tonsil and leukocyte tissues. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV69 is provided in Example 2.

The NOV69 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: inflammatory and immune system-related diseases such as rheumatoid arthritis and inflammatory bowel disease, periodontitis, hypothyroidism, congenital, due to thyroid

dysgenesis or hypoplasia; osteoarthritis of distal interphalangeal joints; selective T-cell defect; nephronophthisis, juvenile; purpura fulminans, neonatal; susceptibility to infections such as viral and bacterial; thrombophilia due to protein C deficiency; as well as other diseases, disorders and conditions.

5 The NOV69 nucleic acid encoding the IL1-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a Interleukin 1 epsilon-like protein includes the nucleic acid whose sequence is provided in Table 69A or 69C, or a fragment thereof. The invention also includes a mutant or
10 variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 69A or 69C while still encoding a protein that maintains its Interleukin 1 epsilon-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 69A or 69C, including nucleic acid fragments that are complementary to any of the nucleic
15 acids just described.

 The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least
20 in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 2% of the bases may be so changed.

 These materials are further useful in the generation of antibodies that bind
25 immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV70

30 NOV70 includes two OS9-like proteins, designated herein as NOV70a and NOV70b.

NOV70a

The disclosed NOV70a (alternatively referred to herein as CG56878-01) includes the 2739 nucleotide sequence (SEQ ID NO:241) shown in Table 70A. A NOV70a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 86-88 and ends with a stop codon at nucleotides 2090-2092. The disclosed NOV70a maps to human chromosome 12q13.

Table 70A. NOV70a Nucleotide Sequence (SEQ ID NO:241)

```
TTGCACTCTCCACACCCCTTTTCTTTTCGTCCGCTCTTCGCTTATTTCTCCCGCCGTCTC
CTCTGCATAAGAAGGGGAACGAAAGATGGCGGCGGAAACGCTGCTGTCCAGTTTGTTAGG
ACTGCTGCTTCTGGGACTCCTGTTACCCGCAAGTCTGACCGGCGGTGTCCGGAGCCTGAA
CCTGGAGGAGCTGAGTGAGATGCGTTATGGGATCGAGATCCTGCCGTTGCCGTGTCATGGG
AGGGCAGAGCCAATCTCGGACGTGGTGATTGTCTCCTTAAGTACAAACAGCGCTATGA
GTGTCGCCTGCCAGCTGGAGCTATTCACTTCCAGCGTGAAAGGGAGGAGAAACACCTGC
TTACCAAGGGCCTGGGATCCCTGAGTTGTTGAGCCCAATGAGAGATGCTCCCTGCTTGCT
GAAGACAAAGGACTGGTGACATATGAATTCTGTTATGGACGCCACATCCAGCAATACCA
CATGGAAGATTCAAGATCAAAGGTGAAGTCTCTATCTCGGCTACTACCAATCAGCCTT
CGACTGGGATGATGAAACAGCCAAGGCTCCAAGCAGCATCGTCTTAAACGCTACCACAG
CCAGACCTATGGCAATGGGTCCAAGTGCGACCTTAATGGGAGGCCCCGGGAGGCCGAGGT
TCGGTTCCTCTGTGACGAGGGTGCAGGTATCTCTGGGGACTACATCGATCGCGTGGACGA
GCCCTTGCTCTGCTTATGTGCTGACCATTGCGACTCCTCGGCTCTGCCCCCACCCTCT
CCTCCGGCCCCCACCAGTGCTGCACCGCAGGCCATCCTCTGTCAACCTTCCCTACAGCC
TGAGGAGTACATGGCTACGTTTCAGAGGCAAGCCGTAGACTCAAAGCAGTATGGAGATAA
AATCATAGAGGAGCTGCAAGATCTAGGCCCCAAGTGTGGAGTGAGACCAAGTCTGGGGT
GGCACCCTAAAAGATGGCAGGTGCGAGCCCCGACCAAGGATGACAGTAAGGACTCAGATT
CTGGAAGATGCTTAATGAGCCAGAGGACCAGGCCCCAGGAGGGGAGGAGGTGCCGGCTGA
GGAGCAGGACCCAAGCCCTGAGGCAGCAGATTACGCTTCTGGTGCTCCCAATGATTTTCA
GAACAACGTGCAGGTCAAAGTCATTGGAAGCCCTGCGGATTTGATTGATTCATAGAGGA
GCTGAAAGGTGGAACAAAAAGGGGAAGCCAAATATAGGCCAAGAGCAGCTGTGGATGA
TGCTGCAGAAGTCCCTCAGAGGGAACCAAGAGAAGGAAGGGGTGATCCAGAACGGCAGAG
AGAGATGGAAGAAGAGAGGAGATGAGGATGAGGATGAGGATGAAGATGAGGATGAACGGCA
GTTACTGGGAGAATTTGAGAAGGAAGTGAAGGGATCCTGCTTCCGTGACACCGAGACCG
GCTCCGTTCCGAGGTGAAGGCTGGCATGGAGCGGGAACCTGGAGAACATCATCCAGGAGAC
AGAGAAAGAGCTGGACCCAGATGGGCTGAAGAAGGAGTCAGAGCGGGATCGGGCAATGCT
GGCTCTCACATCCACTCTCAACAACTCATCAAAAGACTGGAGGAAAAACAGAGTCCAGA
GCTGTGTAAGAAGCACAAAGAAAAGAGGGTTGTCCCCAAAAAGCCTCCCCCATCACCCCA
ACCTACAGAGGAGGATCCTGAGCACAGAGTCCGGGTCCGGGTACCAAGCTCCGTCTCGG
AGGCCCTAATCAGGATCTGACTGTCTCGAGATGAAACGGGAAAACCCACAGCTGAAACA
AATCGAGGGGCTGGTGAAGGAGCTGCTGGAGAGGGAGGACTCACAGCTGCAGGGAAAAT
TGAGATCAAAAATTGTCCGCCATGGGCTGAAGGACTGAAGAGGGTGCACGTTGGCTGAC
TGATGAGGACACGAGAAACCTCAAGGAGATCTTCTTCAATATCTTGGTGCCGGGAGCTGA
AGAGGCCCAGAAGGAACGCCAGCGGCAGAAAGAGCTGGAGAGCAATTACCGCCGGGTGTG
GGGCTCTCCAGGTGGGAGGGGCACAGGGACCTGGACGAATTTGACTTCTGAGACCAACA
CTACACTTGACCTTACGGAATCCAGACTCTTCTTGGACTGGCTTGCCCTCCTCCACC
TCCCCACCCTGGAACCCCTGAGGGCCAAACAGCAGAGTGGAGCTGAGCTGTGGACCTCTC
GGGCAACTCTGTGGGTGTGGGGGCCCTGGGTGAATGCTGCTGCCCCGTGTCGACGCCACC
TTGAGACCTCACCGGCCCTGTGATATTGCTCTCCTGAACTCTCACTCAATCCTCTTCCT
CTCCTCTGTGGCTTTTCTGTTATTGTCCCCTAATGATAGGATATTCCTGTGCTGCTTACC
TGGAGATTCAAGTAGGATCTTTTGAAGTGGAGGTGGGTAGAGAGAGCAAGGAGGGCAGGACA
CTTAGCAGGCACTGAGCAAGCAGGCCCCCACCTGCCCTTAGTGATGTTTGGAGTCTGTTT
ACCCTCTTCTATTGAATTGCCTTGGGATTTCTTCTCCCTTTCCCTGCCACCCTGTCCC
CTACAATTTGTGCTTCTGAGTTGAGGAGCCTTACCTCTGTTGCTGAGGAAATGGTAGAA
TGCTGCCTATCACCTCCAGCACAATCCAGCGAAAAAGGTGTGAAGCACCCACCATGTTT
TTGAACAATCAGGTTTCTAAATAAACAACCTGGACCATCA
```

A NOV70a polypeptide (SEQ ID NO:242) encoded by SEQ ID NO:241 is 668 amino acids in length and is presented using the one-letter amino acid code in Table 70B. The Psort profile for NOV70a predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.8200. In alternative embodiments, a NOV70a

polypeptide is located in the nucleus with a certainty of 0.1260. The Signal P predicts a likely cleavage site for a NOV70a peptide is between positions 25 and 26, *i.e.*, at the dash in the sequence SLT-GG.

Table 70B. NOV70a Polypeptide Sequence (SEQ ID NO:242)

MAAETLLSSLGLLLLLGLLLPASLTGGVGSINLEELSEMRYGIEILPLPVMGGQSQSSDV
VIVSSKYKQRYECLRPAGAIHFQREEREETPAYQGPPIPELLSPMRDAPCLLKTKDWWTY
EFCYGRHIQQYHMEDESEIKGEVLYLGYYSQAFDWDDETAKASKQHRLKRYHSQTYGNGSK
CDLNGRPREAEEVFLCDEGAGISGDYIDRVDEPLSCSYVLTIRTPRLCPHPLLRPPPSAA
PQAILCHPSLQPEEYMAVVRQAVDSKOYGDKIIIEELQDLGPQVWSETKSGVAPQKMAGA
SPTKDDSKDSDFWKMLNEPEDQAPGGEVPAEEQDPSPEAADSASGAPNDFQNNVQVKVI
RSPADLIRFIEELKGGTKKGKPNIGQEOPVDDAAEVPOREPEKERGDPERQREMEEEDE
DEDEDEDEDEROLLGEFEKELEGILLPSDRDLRSEVKAGMERELENIIQETEKELDPDG
LKKESERDRAMLAALTSTLNKLIKRLKQSPKELVKKHKKRVVPPKPPSPQPTTEEDPEH
RVRVRVTKLRLGGPNQDLTVLEMKRENPLKQIEGLVKELLEREGLTAAGKIEIKIVRPW
AEGTEEGARWLTDEDTRNLKEIFFNILLVPGAEAAQKERQROKELESNYRRVWGSPPGEGT
GDLDEFDF

NOV70b

The disclosed NOV70b (alternatively referred to herein as CG56868-04) includes the 2702 nucleotide sequence (SEQ ID NO:) shown in Table 70C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 86-88 and ends with a TGA codon at nucleotides 2036-2038. The disclosed NOV70b maps to human chromosome 12q13.

Table 70C. NOV70b Nucleotide Sequence (SEQ ID NO:243)

TTGCACTCTCCACACCCCTTTTCTTTTCGTCCGCTCTTCGCTTATTTCTCCGCGCTCTC
CTCTGCATAAGAAAGGGGAACGAAAGATGGCGGCGGAAACGCTGCTGTCCAGTTTGTAGG
ACTGCTGCTTCTGGGACTCCTGTTACCCGCAAGTCTGACCGGCGGTGTCCGGAGCTGAA
CCTGGAGGAGCTGAGTGAGATGCGTTATGGGATCGAGATCCTGCCGTTGCCTGTCTATGG
AGGGCAGAGCCAATCTTCGACGCTGGTGATTGTCTCCTCTAAGTACAAACAGCGCTATGA
GTGTGCGCTGCGAGCTGGAGCTATCACTTCCAGCGTGAAAGGGAGGAGGAAACACCTGC
TTACCAAGGGCCTGGGATCCCTGAGTTGTTGAGCCCAATGAGAGATGCTCCCTGTCTGCT
GAAGACAAAGGACTGGTGACATATGAATTCTGTTATGGACGCCACATCCAGCAATACCA
CATGGAAGATTGAGAGATCAAAGGTGAAGTCCCTCTATCTCGGCTACTACCAATCAGCCTT
CGACTGGGATGATGAAACAGCCAAGGCTCCAAGCAGCATCGTCTTAAACGCTACCAACAG
CCAGACCTATGGCAATGGGTCCAAGTGCGACCTTAATGGGAGGCCCGGGAGGCGGAGGT
TCGGTTCTCTGTGACGAGGGTGACGGTATCTCTGGGACTACATCGATCGCGTGGACGA
GCCCTTGCTCTCTTATGTGCTGACCATTCGCACTCCTCGGCTCTGCCCCACCTCT
CCTCCGGCCCCACCCAGTGTGTCACCAAGGCGATCCTCTGTCAACCCTTCCCTACAGCC
TGAGGAGTACATGGCCTACGTTTCAAGGCAAGCCGACTCAAAGCAGTATGGAGATAAAAT
CATAGAGGAGCTGCAAGATCTAGGCCCCCAAGTGTGGAGTGAGACCAAGTCTGGGGTGGC
ACCCCAAAAGATGGCAGGTGCGAGCCCGACCAAGGATGACAGTAAGGACTCAGATTCTTG
GAAGATGCTTAATGAGCCAGAGGACCAAGCCCCAGGAGGGAGGAGTGCCTGCTGAGGA
GCAGGACCAAGCCCTGAGGCAGCAGATTGAGCTTCTGGTGTCTCCCAATGATTTTCAGAA
CAACGTGCAGGTCAAAGTCATTGCAAGCCCTGCGGATTTGATTGATTGATTGAGGAGCT
GAAAGGTGGAACAAAAAAGGGGAAGCCAAATATAGGCCAAGAGCAGCCTGTGGATGATGC
TGCAGAAGTCCCTCAGAGGGAACCAAGAGAAGGAAGGGGTGATCCAGAACGGCAGAGAGA
GATGGAAGAAGAGGAGGATGAGGATGAGGATAAGATGAGGATGAACGGCAGTTACTGGGG
AGAATTTGAGAAGGAAGTGAAGGGATCCTGCTTCCGTGACAGCCGAGACCGGCTCCGTTT
GGAGACAGAGAAAGAGCTGGAACCCAGATGGGCTGAAGAAGGAGTCAGAGCGGGATCGGGC
AATGCTGGCTCTCACATCCACTCTCAACAACTCATCAAAGACTGGAGGAAAAACAGAG
TCCAGAGCTGGTGAAGAAGCACAAGAAAAAGAGGGTTGTCCCCAAAAAGCCTCCCCATC
ACCCCAACCTACAGAGGAGGATCCTGAGCACAGAGTCCGGGTCCGGGTACCAAGCTCCG

Sequences producing High-scoring Segment Pairs:	Score	P (N)
patp:AAG76089 Human colon cancer antigen protein	423	3.4e-52
patp:AAG41826 Arabidopsis thaliana protein fragment	298	2.6e-25
patp:AAG41827 Arabidopsis thaliana protein fragment	279	2.9e-23
patp:AAG41828 Arabidopsis thaliana protein fragment	279	2.9e-23
patp:AAU32255 Novel human secreted protein #2746	232	3.4e-18

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 2735 of 2739 bases (99%) identical to a gb:GENBANK-ID:HSU41635|acc:U41635.1 mRNA from *Homo sapiens* (Human OS-9 precursor mRNA, complete cds). The full amino acid sequence of the protein of the invention was found to have 667 of 668 amino acid residues (99%) identical to, and 667 of 668 amino acid residues (99%) similar to, the 667 amino acid residue ptnr:SWISSPROT-ACC:Q13438 protein from *Homo sapiens* (Human) (PROTEIN OS-9 PRECURSOR). The sequence of this invention has 1 amino acid inserion, compared to ptnr:SWISSPROT-ACC:Q13438 protein from *Homo sapiens* (Human) (PROTEIN OS-9 PRECURSOR). NOV70 also has homology to the other proteins shown in the BLASTP data in Table 70F.

Table 70F. NOV70 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 5803109 ref NP_006803.1 (NM_006812)	amplified in osteosarcoma [<i>Homo sapiens</i>]	667	667/668 (99)	667/668 (99)	0.0
gi 12653521 gb AAH00532.1 AAH00532 (BC000532)	Similar to amplified in osteosarcoma [<i>Homo sapiens</i>]	612	512/513 (99)	512/513 (99)	0.0
gi 17986213 gb AAC39523.2 (U81031)	OS9 [<i>Homo sapiens</i>]	474	474/475 (99)	474/475 (99)	0.0
gi 13905114 gb AAH06844.1 AAH06844 (BC006844)	Unknown (protein for IMAGE:3598453) [<i>Mus musculus</i>]	571	435/630 (69)	495/630 (78)	0.0
gi 13676348 gb AAH06506.1 AAH06506 (BC006506)	Similar to amplified in osteosarcoma [<i>Homo sapiens</i>]	376	373/374 (99)	374/374 (99)	e-161

This BLASTP data is displayed graphically in the ClustalW in Table 70G. A multiple sequence alignment is given, with the NOV70a and b proteins being shown on lines 1 and 2 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 70F.

Table 70G. ClustalW Alignment of NOV70

NOV70a	(SEQ ID NO:242)	
NOV70b	(SEQ ID NO:244)	
gi 5803109	(SEQ ID NO:649)	
gi 12653521	(SEQ ID NO:650)	
gi 17986213	(SEQ ID NO:651)	
gi 13905114	(SEQ ID NO:652)	
gi 13676348	(SEQ ID NO:653)	

		10	20	30	40	50	60	
NOV70a	1	MAAETLLSLLG	LLGLLLPASLTGGVGS	LNLEELSEMRYG	IEILPLPVMGGQ	SSQSSDV	60	
NOV70b	1	MAAETLLSLLG	LLGLLLPASLTGGVGS	LNLEELSEMRYG	IEILPLPVMGGQ	SSQSSDV	60	
gi 5803109	1	MAAETLLSLLG	LLGLLLPASLTGGVGS	LNLEELSEMRYG	IEILPLPVMGGQ	SSQSSDV	60	
gi 12653521	1	MAAETLLSLLG	LLGLLLPASLTGGVGS	LNLEELSEMRYG	IEILPLPVMGGQ	SSQSSDV	60	
gi 17986213	1	-----	-----	-----	-----	-----	1	
gi 13905114	1	-----	-----	-----	-----	PLPVMGGQSSQSSDV	14	
gi 13676348	1	-----	-----	-----	-----	-----	1	

		70	80	90	100	110	120	
NOV70a	61	VIIVSSKYKQRYE	CRLPAGAIHFQRE	REEETPAYQGPGI	PELLSPMRDAP	CLLTKTDWWTY	120	
NOV70b	61	VIIVSSKYKQRYE	CRLPAGAIHFQRE	REEETPAYQGPGI	PELLSPMRDAP	CLLTKTDWWTY	120	
gi 5803109	61	VIIVSSKYKQRYE	CRLPAGAIHFQRE	REEETPAYQGPGI	PELLSPMRDAP	CLLTKTDWWTY	120	
gi 12653521	61	VIIVSSKYKQRYE	CRLPAGAIHFQRE	REEETPAYQGPGI	PELLSPMRDAP	CLLTKTDWWTY	120	
gi 17986213	1	-----	-----	-----	-----	-----	1	
gi 13905114	15	VIIVSSKYKQRYE	CRLPAGAIHFQRE	REEETPAYQGPGI	PELLSPMRDAP	CLLTKTDWWTY	74	
gi 13676348	1	-----	-----	-----	-----	-----	1	

		130	140	150	160	170	180	
NOV70a	121	EFCYGRHIQQYH	MEDSEIKGEVLYL	GYQSADFDD	DETAKASKQHRL	KRYHSQTYGNGSK	180	
NOV70b	121	EFCYGRHIQQYH	MEDSEIKGEVLYL	GYQSADFDD	DETAKASKQHRL	KRYHSQTYGNGSK	180	
gi 5803109	121	EFCYGRHIQQYH	MEDSEIKGEVLYL	GYQSADFDD	DETAKASKQHRL	KRYHSQTYGNGSK	180	
gi 12653521	121	EFCYGRHIQQYH	MEDSEIKGEVLYL	GYQSADFDD	DETAKASKQHRL	KRYHSQTYGNGSK	180	
gi 17986213	1	-----	-----	-----	-----	-----	1	
gi 13905114	75	EFCYGRHIQQYH	MEDSEIKGEVLYL	GYQSADFDD	DETAKASKQHRL	KRYHSQTYGNGSK	134	
gi 13676348	1	-----	-----	-----	-----	-----	1	

		190	200	210	220	230	240	
NOV70a	181	CDLNGCPREAEV	RFLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	240	
NOV70b	181	CDLNGCPREAEV	RFLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	240	
gi 5803109	181	CDLNGCPREAEV	RFLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	240	
gi 12653521	181	CDLNGCPREAEV	RFLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	240	
gi 17986213	1	-----	FLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	47	
gi 13905114	135	CDLNGCPREAEV	RFLCDEGAGISGDY	IDRVDEPLSCSY	VLTIPTRLCPH	PLLRPPPSAA	194	
gi 13676348	1	-----	-----	-----	-----	-----	1	

		250	260	270	280	290	300	
NOV70a	241	PQAILCHPSLQPEE	YMAVVRQAVD	SKQYGDKIIIEEL	QDLGPQVWSETK	SGVAPQKMAGA	300	
NOV70b	241	PQAILCHPSLQPEE	YMAVVRQAVD	SKQYGDKIIIEEL	QDLGPQVWSETK	SGVAPQKMAGA	299	
gi 5803109	241	PQAILCHPSLQPEE	YMAVVRQAVD	SKQYGDKIIIEEL	QDLGPQVWSETK	SGVAPQKMAGA	299	
gi 12653521	241	PQAILCHPSLQPEE	YMAVVRQAVD	SKQYGDKIIIEEL	QDLGPQVWSETK	SGVAPQKMAGA	299	
gi 17986213	48	PQAILCHPSLQPEE	YMAVVRQAVD	SKQYGDKIIIEEL	QDLGPQVWSETK	SGVAPQKMAGA	106	
gi 13905114	195	PQAILCHPALQPEE	YMAVLRQAE	SKQHEEKTTEE	VQDTDRQVWSG	SKAGAPPKEDV	253	
gi 13676348	1	-----	-----	-----	-----	HEEKMAGA	8	

		310	320	330	340	350	360	
NOV70a	301	SPTKDDSKDSDFW	KMLNEPEDQAPG	GEEVPAEEQDP	SPEAADSASGAP	NDFQNNVQVKV	359	
NOV70b	300	SPTKDDSKDSDFW	KMLNEPEDQAPG	GEEVPAEEQDP	SPEAADSASGAP	NDFQNNVQVKV	358	
gi 5803109	300	SPTKDDSKDSDFW	KMLNEPEDQAPG	GEEVPAEEQDP	SPEAADSASGAP	NDFQNNVQVKV	358	
gi 12653521	300	SPTKDDSKDSDFW	KMLNEPEDQAPG	GEEVPAEEQDP	SPEAADSASGAP	NDFQNNVQVKV	358	
gi 17986213	107	SPTKDDSKDSDFW	KMLNEPEDQAPG	GEEVPAEEQDP	SPEAADSASGAP	NDFQNNVQVKV	165	
gi 13905114	254	SPAKEEKESELN	KIQGPEEQAAARE	EAQAGEQDLN	HEAAADAPSP	PNDQNNVQVKV	311	

gi 13676348	9	SPTKDDSKDSDFWKMLNEPEDQAPGGEEVPAEEQDPSPSE	AADSASGAPNDFQNNVQVKV	67			
		370	380	390	400	410	420
NOV70a	360	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	416
NOV70b	359	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	415
gi 5803109	359	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	415
gi 12653521	359	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	415
gi 17986213	166	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	222
gi 13905114	312	IRSPADLIRFIEELK-AAEKGKPSVRREQPCDDTTEAPQREAE	CTKAKCKGCEPPGLME				370
gi 13676348	68	IRSPADLIRFIEELKGGTKKGKPNIGQEQPVDDAAEVPQREPE	-----	KERGD	P	ERQREME	124
		430	440	450	460	470	480
NOV70a	417	E-----EDEDEDEDEDEDERQLLGEFEKELEGILLPSDRDRLRSEVKAGMERELENIIQET					472
NOV70b	416	E-----EDEDEDKMRMNG--SYWGEFEKELEGILLPSDRDRLR-----SET					454
gi 5803109	416	E-----EDEDEDEDEDEDERQLLGEFEKELEGILLPSDRDRLRSEVKAGMERELENIIQET					471
gi 12653521	416	E-----EDEDEDEDEDEDERQLLGEFEKELEGILLPSDRDRLRSEVKAGMERELENIIQET					471
gi 17986213	223	E-----EDEDEDEDEDEDERQLLGEFEKELEGILLPSDRDRLRSEVKAGMERELENIIQET					278
gi 13905114	371	EDGDDDEEEEEEDEDEEEQLLGEFEKELEGILLPSDRERLRSEVKAGMERELENIIQET					430
gi 13676348	125	E-----EDEDEDEDEDEDERQLLGEFEKELEGILLPSDRDRLRSEVKAGMERELENIIQET					180
		490	500	510	520	530	540
NOV70a	473	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					532
NOV70b	455	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					514
gi 5803109	472	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					531
gi 12653521	472	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					531
gi 17986213	279	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					338
gi 13905114	431	EKELDPDGLKKESEREQATLALTSTLDKLIKRLEENQSPELVCKYKKRVRVPCKPPSPH					490
gi 13676348	181	EKELDPDGLKKESERDRAMLALTSTLNKLIKRLEEKQSPELVKKHKKKRVVPKKPPSPQ					240
		550	560	570	580	590	600
NOV70a	533	PTEEDPEHRVRVRVTKLRLGGPNQDLTVLEMKRENPKQKQIEGLVKELLEREGTLAAGKI					592
NOV70b	515	PTEEDPEHRVRVRVTKLRLGGPNQDLTVLEMKRENPKQKQIEGLVKELLEREGTLAAGKI					574
gi 5803109	532	PTEEDPEHRVRVRVTKLRLGGPNQDLTVLEMKRENPKQKQIEGLVKELLEREGTLAAGKI					591
gi 12653521	532	PT-----GKI					536
gi 17986213	339	PTEEDPEHRVRVRVTKLRLGGPNQDLTVLEMKRENPKQKQIEGLVKELLEREGTLAAGKI					398
gi 13905114	491	PT-----GKI					495
gi 13676348	241	PTEEDPEHRVRVRVTKLRLGGPNQDLTVLEMKRENPKQKQIEGLVKELLEREGTLAAGKI					300
		610	620	630	640	650	660
NOV70a	593	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					652
NOV70b	575	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					634
gi 5803109	592	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					651
gi 12653521	537	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					596
gi 17986213	399	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					458
gi 13905114	496	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					555
gi 13676348	301	EIKIVRPWAEGTTEEGARWLTDEDTRNLKEIFFNILVPGAEAAQKERQROKELESNYRRVW					360
		670					
NOV70a	653	GSPGGEGTGDLDEFDF					668
NOV70b	635	GSPGGEGTGDLDEFDF					650
gi 5803109	652	GSPGGEGTGDLDEFDF					667
gi 12653521	597	GSPGGEGTGDLDEFDF					612
gi 17986213	459	GSPGGEGTGDLDEFDF					474
gi 13905114	556	GSPGGEGTGDLDEFDF					571
gi 13676348	361	GSPGGEGTGDLDEFDF					376

Amplification and overexpression of genes involved in cellular growth control occur frequently in human tumors. Using a chromosome microdissection-based hybrid-selection

strategy, the OS-9 gene has been identified within 12q13-15, a region frequently amplified in human cancers. The full-length OS-9 cDNA sequence consists of 2785 bp from which an open reading frame (ORF) with 667 amino-acid residues has been deduced. The predicted polypeptide is water soluble and acidic. The OS-9 gene encodes a 2.8-kb mRNA transcribed in all 16 human tissues examined, suggesting that OS-9 is ubiquitously expressed in human tissues. OS-9 is co-amplified with CDK4 in three of five sarcoma tissues. Homology analysis of the amino-acid sequence has revealed significant similarities between OS-9 and two ORFs deduced from genomic sequences in *Caenorhabditis elegans* and *Saccharomyces cerevisiae*. The region of similarity extends over 200 residues (approximately one-third of each ORF), and eight cysteines were conserved in all three ORFs. These observations suggest that this region comprises a functional domain present in a novel evolutionarily conserved gene family defined by OS-9.

The OS-9 genomic DNA has been isolated and characterized from a human BAC library. Sequencing of the genomic DNA has shown that the gene spanned approximately 30.4 kbp and had 15 exons. The 1,010 bp sequence of the 5' upstream region has also been determined. The potential binding-sequence motifs TATA and CCAAT for general transcription factors have been found in the 5' upstream region. Primer extension analysis has revealed two putative transcription start sites.

Three isoforms of OS-9 cDNA have been isolated from a myeloid leukemia HL-60 cDNA library and characterized. Isoform 1 consisted of 2,700 bp, from which a 667 amino acid sequence was deduced and found to be identical with that of OS-9 cDNA from osteosarcoma cells. Isoform 2 cDNA lacked a 165 nucleotide sequence in the coding region. Isoform 3 cDNA had an additional 45 bp deletion in the coding region. Isoforms 2 and 3 encode 612 and 597 amino acid polypeptides, respectively. Comparison of their cDNA sequences with the genomic structure has indicated that three isoforms are splice variants. Reverse transcription-polymerase chain reaction analysis has shown predominant expression of isoform 2 mRNA in myeloid leukemia HL-60 cells, osteosarcoma Osa-CL cells and rhabdomyosarcoma Rh30 cells. Northern blotting has revealed similar levels of expression of OS-9 gene in various tumor cell lines of sarcoma cells, carcinoma cells and myeloid leukemia cells, but 3-4 times higher expression in Osa-CL cells and Rh30 cells containing a homogeneously staining region of 12q13-15. OS-9 expression decreased in differentiation-induced HL-60 cells. The above data suggests a possible involvement of OS-9 in cell growth and tumour development.

The NOV70 disclosed in this invention is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, amnion, aorta, ascending colon, bone, bronchus, cartilage, cervix, colon, cornea, coronary artery, dermis, duodenum, epidermis, epididymis, hair follicles, hypothalamus, islets of langerhans, kidney cortex, liver, lung, lymph node, lymphoid tissue, esophagus, ovary, parathyroid gland, peripheral blood, pineal gland, respiratory bronchiole, retina, skin, thymus, tonsils, umbilical vein, urinary bladder, vulva, whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV70 is provided in Example 2.

The NOV70 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention may have efficacy for the treatment of patients suffering from leukemias, sarcomas and other types of cancer, as well as other diseases, disorders and conditions. OS-9 was co-amplified with CDK4 in three of five sarcoma tissues (Mol Carcinog 1996 Apr;15(4):270-5). Three isoforms of OS-9 cDNA were found in a myeloid leukemia HL-60 cDNA library and reverse transcription-polymerase chain reaction analysis has shown predominant expression of isoform 2 mRNA in myeloid leukemia HL-60 cells, osteosarcoma OsA-CL cells and rhabdomyosarcoma Rh30 cells. Northern blotting has revealed similar levels of expression of OS-9 gene in various tumor cell lines of sarcoma cells, carcinoma cells and myeloid leukemia cells (J Biochem (Tokyo) 1998 May;123(5):876-82).

The NOV70 nucleic acid encoding the OS-9-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a OS-9-like protein includes the nucleic acid whose sequence is provided in 70A or 70C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in 70A or 70C while still encoding a protein that maintains its OS-9-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence of 70A or 70C, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

5 Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to
10 about 1% of the bases may be so changed.

The novel protein of the invention includes the OS-9-like protein whose sequence is provided in Table 70B or 70D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 70B or 70D while still encoding a protein that maintains its OS-9-like activities and physiological
15 functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using
20 prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV71

The disclosed NOV71 (alternatively referred to herein as CG56906-01) includes the
25 2081 nucleotide sequence (SEQ ID NO:245) shown in Table 71A. A NOV71 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 55-57 and ends with a stop codon at nucleotides 1978-1980. The disclosed NOV71 maps to human chromosome Xp11.

Table 71A. NOV71 Nucleotide Sequence (SEQ ID NO:245)

```
GCGGCCGCGGCCTCGGCCTCCTCTGCGGCGCGCCGAGGACAGCAGCGCCATGGAG
GAGCTCGTACTGAGAAGGAGGCGGAGGAGGCCACCGGCAAGACAGCGTGAGCCTGCTC
ACCTTCATCCTGCTGCTCAGCTCACCATCCTCACCATCTGGCTCTTCAAGCACCGCCGG
GTGCGCTTCTGCACGAGACCGGGCTGGCCATGATCTATGGGCTCATCGTTGGGGTGATC
CTGAGGTATGGTACCCCTGCTACCAAGTGGCCGTGACAAATCACTCAGCTGCACTCAGGAA
GACAGGGCCTTCAGTACCTTATTAGTGAATGTCAGCGGAAAGTTCTTCAATACACTCTG
AAAGGAGAAATCAGTCTGGCAAGATCAACAGCGTAGAGCAGAATGATATGTACGGAAG
GTAACATTCGATCCAGAAGTATTTTCAACATTCTTCTGCCTCCAATTATTTTCATGCT
GGATACAGCTTAAAGAAGAGACACTTTTTCAGAAATCTTGGATCTATACTGGCCTATGCC
TTCTTGGGACTGCTGTTTCATGCTTCATTATTGGAAATCTCATGTATGGTGTGGTGAAG
```

CTCATGAAGATTATGGGACAGCTCTCAGATAAATTTTACTACACAGATTGTCTCTTTTTT
GGAGCAATCATCTCTGCCACTGACCCAGTACTGTGCTGGCGATATTTAATGAATTGCAT
GCAGACGTGGATCTTTTACGCACCTCTTTTTGGAGAGAGCGTCTCTAAATGATGCTGTTGCC
ATTGTAAGTCTCTCGTCTATTGTTGCCTACCAGCCAGCGGGACTGAACACTCAGCCCTTT
GATGCTGCTGCCTTTTTTAAGTCAGTTGGCATTTTTCTAGGTATATTTAGTGGCTCTTTT
ACCATGGGAGCTGTGACTGGTGTGTGACTGCTCTAATATCCTTTTTTACAGAATGCCAAC
GTGACTAAGTTTACCAAAGTGCCTGCTTCCCCCTGCTGGAGACGGCGCTGTTCTTCTCC
ATGCTCTGGAGCAGCTTCTCTTGGCAGAAGCCTGCGGATTTACAGGTGTTGTAGCTGTC
CTTTTCTGTGGAATCACACAAGCTCATTACACCTACAACAATCTGTGCGTGGGAATCAAGA
AGTCGAACCAAGCAGCTCTTTGAGGTGTTACATTTCTTGGCAGAGAACTTCATCTTCTCC
TACATGGGCCTGGCAGCTGTTTACCTTCCAGAGCAGCTTTTACAGCCCATTTTTCATCATC
GGAGCTTTTGTGCGCATCTTCTTGGGAGAGCCGCGCACATCTACCCGCTCTCTCTTCTTC
CTCAACTTGGGAGAGGAGCATAAGATTGGCTGGAATTTTCAACACATGATGATGTTTTC
GGCCTCAGGGGAGCAATGGCATTGCGTGGCCATCCGTGACACGGCATCTATGCTCGC
CAGATGATGTTACAGACACCCCTTCTCATTTGTTCTTCACTGTCTGGATCATTTGGAGGA
GGCAGACACCCATGTTGTCTAGGCTTAACATCAGGTTGGACGGCCAGATTGATGTTTTC
GGAAACCGGACAAAACAGGAGAGCGCATGGATATTCAGGCTGTGGTACAGCTTTGATCAC
AATTACCTGAAGCCCATCTCACACAGTGGTCCCCCACTAACCACCACGCTCCCCGCC
TGGTGTGGCTTACTAGTCTGATGTCTGACAGTCCCCAGGTGTACGATAACCAAGAGCCA
CTGAGAGAGGAAGACTCTGATTTTCATCTGACCGAAGGCGACCTGACATTGACCTACGGG
GACAGCACAGTGAAGTGGCTCCTCAAGTTCGCACACCGCTCCACGAGTCTGGAG
GGCAGCCGGAGAACGAAGAGCAGCTCGGAGGAAGTGTGGAGCGAGACCTGGGAATGGGA
GACCAGAAGGTTTCGAGCCGGGACCCGCTAGTGTTCCTTGGGAAGATAATGCTTGA
CTTTCCCCCAAGCCCTGGCGCATGGGTAGGCTCCCGATGGGGTGAGGACAGCTGCAA
GCCCTAGTGTGTTGGAGGTGGGCGAGTGAAGTGAAC

A NOV71 polypeptide (SEQ ID NO:246) encoded by SEQ ID NO:245 is 641 amino acids in length and is presented using the one-letter amino acid code in Table 71B. The Psort profile for NOV71 predicts that this sequence is likely to be a Type III 6 membrane protein, has a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.8200. In alternative embodiments, a NOV71 polypeptide is located to Golgibodies with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.6850. The Signal P predicts a likely cleavage site for a NOV71 peptide is between positions 37 and 38, *i.e.*, at the dash in the sequence IWL-FK.

Table 71B. NOV71 Polypeptide Sequence (SEQ ID NO:246)

MEELATEKEAEESHQRQSVSLTLTILLTLTILTIWLFKHRRVRFLHETGLAMIYGLIVG
VILRYGTPATSGRDKSLSTQEDRAFSTLLVNVSGKF FEYTLKGEISPGKINSVEQNDML
RKVTFDPEVFFNILLPPIIFHAGYSLKKRHFFRNLSILAYAF LGTAVSCFIIGNLMYGV
VKLMKIMGQLSDKFYYTDCLFFGAIISATDPVTVLAI FNEHADVLDLYALLFGESVLNDA
VAIVLSSIVAYQPAGLNTHAFDAAFFKSVGIFLGIFSGSFTMGAVTGVV TALISFLQN
ANVTKFTKLHCFPLETALFFLMSWSTFLAEACGFTGVVAVLF CGITQAHYTYNNLSVE
SRSRKQLFEVLHFLAENFISYMLGALFTFQKHVFSPIFIICAFVAIFLGRAAHYPLS
FFLNLRHRRHKIGWNFQHMMMFSGLRGAMAFALAI RDTASYARQMMFTTLLIVFFTVWII
GGGTTPMLSWLNIRLDGPDSARGNRTKQESAWIFRLWYSFDHNYLKPILTHSGPPLTTTL
PAWCGLLARCLTSPQVYDNQEPLEEDSDFILTEGDLTLTYGDSTVTANGSSSSHTASTS
LEGSRRTKSSSEEVLERDLGMDQKVSSRGTRLVFPLEDNA

A BLAST analysis of NOV71 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV71 had high homology to other proteins as shown in Table 71C.

Table 71C. BLASTX results from PatP database for NOV71

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAB90555 Human secreted protein,	2410	5.1e-250
patp:AAB90637 Human secreted protein,	2410	5.1e-250
patp:AAU02883 Human HsNHE-6 polypeptide - <i>Homo sapiens</i> ,	1693	5.2e-201
patp:AAB90590 Human secreted protein,	902	2.1e-169
patp:AAB90591 Human secreted protein,	902	2.1e-169

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 747 of 1121 bases (66%) identical to a gb:GENBANK-ID:AF030409|acc:AF030409.1 mRNA from *Homo sapiens* (sodium-hydrogen exchanger 6 (NHE-6) mRNA, nuclear gene encoding mitochondrial protein). The full amino acid sequence of the protein of the invention was found to have 391 of 518 amino acid residues (75%) identical to, and 443 of 518 amino acid residues (85%) similar to, the 669 amino acid residue ptnr:SWISSNEW-ACC:Q92581 protein from *Homo sapiens* (Human) (SODIUM/HYDROGEN EXCHANGER 6 (NA(+)/H(+)) EXCHANGER 6 (NHE-6)).

NOV71 also has homology to the other proteins shown in the BLASTP data in Table 71D.

Table 71D. NOV71 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14211919 ref NP_115980.1 (NM_032591)	solute carrier family 9 (sodium/hydrogen exchanger), isoform 7; nonselective sodium potassium/proton exchanger [<i>Homo sapiens</i>]	725	631/681 (92)	632/681 (92)	0.0
gi 3319946 emb CAA18155.1 (AL022165)	dJ71L16.5 (KIAA0267 LIKE putative Na(+)/H(+) exchanger) [<i>Homo sapiens</i>]	616	575/625 (92)	576/625 (92)	0.0
gi 1665827 dbj BAA13449.1 (D87743)	Similar to Human Na+/H+ exchanger 2 (A57644) [<i>Homo sapiens</i>]	666	451/649 (69)	513/649 (78)	0.0
gi 5454070 ref NP_006350.1 (NM_006359)	solute carrier family 9 (sodium/hydrogen exchanger), isoform 6 [<i>Homo sapiens</i>]	669	451/649 (69)	513/649 (78)	0.0
gi 17474970 ref XP_062645.1 (XM_062645)	similar to solute carrier family 9 (sodium/hydrogen exchanger), isoform 7; nonselective sodium potassium/proton exchanger (H. sapiens) [<i>Homo sapiens</i>]	485	292/412 (70)	305/412 (73)	e-148

This BLASTP data is displayed graphically in the ClustalW in Table 71E. A multiple sequence alignment is given, with the NOV71 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 71D.

5

Table 71E. ClustalW Alignment of NOV71									
NOV71	(SEQ ID NO:246)								
gi 14211919	(SEQ ID NO:654)								
gi 3319946	(SEQ ID NO:655)								
gi 1665827	(SEQ ID NO:656)								
gi 5454070	(SEQ ID NO:657)								
gi 17474970	(SEQ ID NO:658)								
<div> <div>102030405060</div> <div> <div>NOV71</div> <div>gi 14211919 </div> <div>gi 3319946 </div> <div>gi 1665827 </div> <div>gi 5454070 </div> <div>gi 17474970 </div> </div> <div> <div>1</div> <div>1</div> <div>1</div> <div>1</div> <div>1</div> </div> <div> <div>MEEPGDAARPGSGRATGAPPPRLLLPLLLGWGLRVAAAASASSSGAAAEDSSAMEEGLATE</div> <div>---</div> <div>MARRGWRRAPLRRGVGSSPRARRLMRPLWLLLA</div> <div>---</div> <div>MEELVTE</div> </div> <div> <div>7</div> <div>60</div> <div>1</div> <div>57</div> <div>60</div> <div>7</div> </div> </div>									
<div> <div>708090100110120</div> <div> <div>NOV71</div> <div>gi 14211919 </div> <div>gi 3319946 </div> <div>gi 1665827 </div> <div>gi 5454070 </div> <div>gi 17474970 </div> </div> <div> <div>8</div> <div>61</div> <div>1</div> <div>58</div> <div>61</div> <div>8</div> </div> <div> <div>KEAEESHRODSVSLTIFILLLTILTITWLFKRRVRLHETGLAMIYGLIVGVILRYGT</div> <div>KEAEESHRODSVSLTIFILLLTILTITWLFKRRVRLHETGLAMIYGLIVGVILRYGT</div> <div>---</div> <div>KQAEESHRODSANLLTIFILLLTILTITWLFKRRRRLHETGLAMIYGLIVGLVLRGT</div> <div>KQAEESHRODSANLLTIFILLLTILTITWLFKRRRRLHETGLAMIYGLIVGLVLRGT</div> <div>KEAEESHRODSVSLTIFILLLTILTITWLFKRRVRLHETGLAMICGLIVGVILRYGT</div> </div> <div> <div>67</div> <div>120</div> <div>11</div> <div>117</div> <div>120</div> <div>67</div> </div> </div>									
<div> <div>130140150160170180</div> <div> <div>NOV71</div> <div>gi 14211919 </div> <div>gi 3319946 </div> <div>gi 1665827 </div> <div>gi 5454070 </div> <div>gi 17474970 </div> </div> <div> <div>68</div> <div>121</div> <div>12</div> <div>118</div> <div>121</div> <div>68</div> </div> <div> <div>PATSGRDKSLSCQEDRAFTLLVNVSGKFFETLKEISPGKINSVEQNDMLRKVT FDP</div> <div>PATSGRDKSLSCQEDRAFTLLVNVSGKFFETLKEISPGKINSVEQNDMLRKVT FDP</div> <div>PATSGRDKSLSCQEDRAFTLLVNVSGKFFETLKEISPGKINSVEQNDMLRKVT FDP</div> <div>HVPSDVNNVTLSCEVQSSPTILLVTFDPEVFENILLPPIIFYAGYSLKRRHFR</div> <div>HVPSDVNNVTLSCEVQSSPTILLVTFDPEVFENILLPPIIFYAGYSLKRRHFR</div> <div>ECTRGRDKLLNCTQEDCAFSTLVVDVSGKFFETLKEISPGKINSVKQNDMLG</div> </div> <div> <div>127</div> <div>180</div> <div>71</div> <div>171</div> <div>174</div> <div>121</div> </div> </div>									
<div> <div>190200210220230240</div> <div> <div>NOV71</div> <div>gi 14211919 </div> <div>gi 3319946 </div> <div>gi 1665827 </div> <div>gi 5454070 </div> <div>gi 17474970 </div> </div> <div> <div>128</div> <div>181</div> <div>72</div> <div>171</div> <div>174</div> <div>121</div> </div> <div> <div>EVFFNILLPPIIFHAGYSLKKRHFFRNLSILAYAF LGTAVSCFIIGNLMYGVVVKLMKIM</div> <div>EVFFNILLPPIIFHAGYSLKKRHFFRNLSILAYAF LGTAVSCFIIGNLMYGVVVKLMKIM</div> <div>EVFFNILLPPIIFHAGYSLKKRHFFRNLSILAYAF LGTAVSCFIIGNLMYGVVVKLMKIM</div> <div>---</div> <div>---</div> <div>---</div> </div> <div> <div>187</div> <div>240</div> <div>131</div> <div>205</div> <div>208</div> <div>121</div> </div> </div>									
<div> <div>250260270280290300</div> <div> <div>NOV71</div> <div>gi 14211919 </div> <div>gi 3319946 </div> <div>gi 1665827 </div> <div>gi 5454070 </div> <div>gi 17474970 </div> </div> <div> <div>188</div> <div>241</div> <div>132</div> <div>206</div> <div>209</div> <div>121</div> </div> <div> <div>GQLSDKFYITDCLFAGAIISATDPVTVLAI FHELHADV DLYALLFGESVLNDAVAIVLSS</div> <div>GQLSDKFYITDCLFAGAIISATDPVTVLAI FHELHADV DLYALLFGESVLNDAVAIVLSS</div> <div>GQLSDKFYITDCLFAGAIISATDPVTVLAI FHELHADV DLYALLFGESVLNDAVAIVLSS</div> <div>GQLAGDFYITDCLFAGAIISATDPVTVLAI FHELQVDV ELYALLFGESVLNDAVAIVLSS</div> <div>GQLAGDFYITDCLFAGAIISATDPVTVLAI FHELQVDV ELYALLFGESVLNDAVAIVLSS</div> <div>---</div> </div> <div> <div>247</div> <div>300</div> <div>191</div> <div>265</div> <div>268</div> <div>121</div> </div> </div>									
<div> <div>310320330340350360</div> <div> <div>NOV71</div> </div> <div> <div>248</div> </div> <div> <div>SIVAYQPAGLNT HAFDAAAF FKS VGI FLGIFSGSFTMGAVTGVVTALISFLQANAVTKFT</div> </div> <div> <div>307</div> </div> </div>									

gi	14211919	301	SIVAYQPAGLNTHAFDAAAFKSVGIFLGIFSGSFTMGAVTG-VNAN-----VTKFT	351
gi	3319946	192	SIVAYQPAGLNTHAFDAAAFKSVGIFLGIFSGSFTMGAVTG-VNAN-----VTKFT	242
gi	1665827	266	SIVAYQPAGDNSHTFDVTAMFKSIGIFLGIFSGSFAMGAATGVVTAL-----VTKFT	317
gi	5454070	269	SIVAYQPAGDNSHTFDVTAMFKSIGIFLGIFSGSFAMGAATGVVTAL-----VTKFT	320
gi	17474970	121	-----KSVGIFLGIFSGCFTMGAVTGVVTAL-----VTKFT	152
<div> <div>370380390400410420</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>308</div> <div>352</div> <div>243</div> <div>318</div> <div>321</div> <div>153</div> </div> <div> <div>KLHCFPLLETALFFFLMSWSTFLLAEACGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> <div>KLHCFPLLETALFFFLMSWSTFLLAEACGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> <div>KLHCFPLLETALFFFLMSWSTFLLAEACGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> <div>KLREFOLLETCLFFFLMSWSTFLLAEAWGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> <div>KLREFOLLETCLFFFLMSWSTFLLAEAWGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> <div>KLDLCPPLLETALFFFLMSWSTFLLAEACGFTGVVAVLFCGITQAHYTYNNLSVESRSRTKQ</div> </div> <div> <div>367</div> <div>411</div> <div>302</div> <div>377</div> <div>380</div> <div>212</div> </div> </div>				
<div> <div>430440450460470480</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>368</div> <div>412</div> <div>303</div> <div>378</div> <div>381</div> <div>213</div> </div> <div> <div>LFEVLHFLAENFIFSYMGLALFTFQKHVFSPTFIIGAFVAIFLGRAAHYPLSFFLNLGR</div> <div>LFEVLHFLAENFIFSYMGLALFTFQKHVFSPTFIIGAFVAIFLGRAAHYPLSFFLNLGR</div> <div>LFEVLHFLAENFIFSYMGLALFTFQKHVFSPTFIIGAFVAIFLGRAAHYPLSFFLNLGR</div> <div>LFELLNFLAENFIFSYMGLTLTFTQNHVFNPTFVVGAFAIFLGRAAHYPLSLLNLGR</div> <div>LFELLNFLAENFIFSYMGLTLTFTQNHVFNPTFVVGAFAIFLGRAAHYPLSLLNLGR</div> <div>LFE-----AENFIFSCMILALFTFQKHVFSPTFIIGAFVAIFLGRAAHYPLSFFLSLGR</div> </div> <div> <div>427</div> <div>471</div> <div>362</div> <div>437</div> <div>440</div> <div>267</div> </div> </div>				
<div> <div>490500510520530540</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>428</div> <div>472</div> <div>363</div> <div>438</div> <div>441</div> <div>268</div> </div> <div> <div>RHKIGWNFOHMMMFSGLRGAMAFALAIKDTASYARQMMFTTLLIVFFTVMWIGGGTTTPM</div> <div>RHKIGWNFOHMMMFSGLRGAMAFALAIKDTASYARQMMFTTLLIVFFTVMWIGGGTTTPM</div> <div>RHKIGWNFOHMMMFSGLRGAMAFALAIKDTASYARQMMFTTLLIVFFTVMWIGGGTTTPM</div> <div>RSKIGSNFOHMMMFAGLRGAMAFALAIKDTASYARQMMFTTLLIVFFTVMWIGGGTTTPM</div> <div>RSKIGSNFOHMMMFAGLRGAMAFALAIKDTASYARQMMFTTLLIVFFTVMWIGGGTTTPM</div> <div>RHKIGWNFOHMMMFSGLRGAMAFALAIKDTASYARQMMFTTTPPIVFFTVMWIGGGTTTPM</div> </div> <div> <div>487</div> <div>531</div> <div>422</div> <div>497</div> <div>500</div> <div>327</div> </div> </div>				
<div> <div>550560570580590600</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>488</div> <div>532</div> <div>423</div> <div>498</div> <div>501</div> <div>328</div> </div> <div> <div>LSWLNIRV-----DGPDSARGNRTK</div> <div>LSWLNIRVGVVEEPSEEDQNEHHWQYFRVGVDPDQPPPNNDSSFQVLQGDGPDARGNRTK</div> <div>LSWLNIRVGVVEEPSEEDQNEHHWQYFRVGVDPDQPPPNNDSSFQVLQGDGPDARGNRTK</div> <div>LSCLHIRVG-----VDSDOE-----HLGVPENERRT-TK</div> <div>LSCLHIRVG-----VDSDOE-----HLGVPENERRT-TK</div> <div>LSWLNIRVSIKEPSKEDHNEHHRQYFRVGVDPDQPPPNNDSSFQVLQGDSPDSARGNRTK</div> </div> <div> <div>507</div> <div>591</div> <div>482</div> <div>525</div> <div>528</div> <div>387</div> </div> </div>				
<div> <div>610620630640650660</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>508</div> <div>592</div> <div>483</div> <div>526</div> <div>529</div> <div>388</div> </div> <div> <div>QESAWIFRLWYSFDHNYLKPILTHSGPPLTTTLPAWCGLLARCLTSPQVYDNQOEPLREED</div> <div>QESAWIFRLWYSFDHNYLKPILTHSGPPLTTTLPAWCGLLARCLTSPQVYDNQOEPLREED</div> <div>QESAWIFRLWYSFDHNYLKPILTHSGPPLTTTLPAWCGLLARCLTSPQVYDNQOEPLREED</div> <div>AESAWIFRMWYNFDHNYLKPILTHSGPPLTTTLPAWCGLLARCLTSPQVYDNQOEPLREED</div> <div>AESAWIFRMWYNFDHNYLKPILTHSGPPLTTTLPAWCGLLARCLTSPQVYDNQOEPLREED</div> <div>QESTWIF-----R-RCQVYDNQOEPLREED</div> </div> <div> <div>567</div> <div>651</div> <div>542</div> <div>585</div> <div>588</div> <div>411</div> </div> </div>				
<div> <div>670680690700710720</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>568</div> <div>652</div> <div>543</div> <div>586</div> <div>589</div> <div>412</div> </div> <div> <div>SDFILTEGDLTLTYGDSVTANGSSSSHTASTSLEGSRRTKSSSSEEVLERDLGMGDQKVS</div> <div>SDFILTEGDLTLTYGDSVTANGSSSSHTASTSLEGSRRTKSSSSEEVLERDLGMGDQKVS</div> <div>SDFILTEGDLTLTYGDSVTANGSSSSHTASTSLEGSRRTKSSSSEEVLERDLGMGDQKVS</div> <div>SDLILNDGDISLTYGDSVTNT-----EPATSSAPRRFMGNSSEDALDRELAFGDHEIV</div> <div>SDLILNDGDISLTYGDSVTNT-----EPATSSAPRRFMGNSSEDALDRELAFGDHEIV</div> <div>SDFILTEGDLTLTYGDSVTANGFSGSSHTASTSLEGSRRTKSSSSEEVLERDLGMGDQKVS</div> </div> <div> <div>627</div> <div>711</div> <div>602</div> <div>638</div> <div>641</div> <div>471</div> </div> </div>				
<div> <div>730740</div> <div> <div>NOV71</div> <div>gi 14211919</div> <div>gi 3319946</div> <div>gi 1665827</div> <div>gi 5454070</div> <div>gi 17474970</div> </div> <div> <div>628</div> <div>712</div> <div>603</div> <div>639</div> <div>642</div> <div>472</div> </div> <div> <div>SRGTRLVFPLEDNA-----</div> <div>SRGTRLVFPLEDNA-----</div> <div>SRGTRLVFPLEDNA-----</div> <div>IRGTRLVLPMDDEPPLNLLDNTRHGPA</div> <div>IRGTRLVLPMDDEPPLNLLDNTRHGPA</div> <div>NOGTRLVFPLEDNV-----</div> </div> <div> <div>641</div> <div>725</div> <div>616</div> <div>666</div> <div>669</div> <div>485</div> </div> </div>				

Table 71F lists the domain description from DOMAIN analysis results against NOV71. This indicates that the NOV71 sequence has properties similar to those of other proteins known to contain this domain.

Table 71F. Domain Analysis of NOV71				
<p><u>gnl</u> <u>Pfam</u> <u>pfam00999</u>, Na_H_Exchanger, Sodium/hydrogen exchanger family. Na/H antiporters are key transporters in maintaining the pH of actively metabolizing cells. The molecular mechanisms of antiport are unclear. These antiporters contain 10-12 transmembrane regions (M) at the amino-terminus and a large cytoplasmic region at the carboxyl terminus. The transmembrane regions M3-M12 share identity with other members of the family. The M6 and M7 regions are highly conserved. Thus, this is thought to be the region that is involved in the transport of sodium and hydrogen ions. The cytoplasmic region has little similarity throughout the family. SEQ ID NO:865</p> <p>CD-Length = 400 residues, 88.5% aligned Score = 276 bits (706), Expect = 3e-75</p>				
NOV71:	123	VTFDPEVFFNILLPPIIFHAGYSLKKRHFFRNLSILAYAF LGTAVSCFIIGNLMYGVVK	182	
Sbjct:	46	V D EVFF ILLPPI+F AG L R FRNLGSIL A LG + IG LMY +V	105	
NOV71:	183	LMKIMGQLSDKFYYTDCLFFGAIISATDPVTLAIFNEL-HADVLDYALLFGESVLNDAV	241	
Sbjct:	106	++ + + L FGAI+SATDPV VLA+ EL L L+FGES+LND V	157	
NOV71:	242	AIVLSSSIVAYQPAGLNTAHFADAAFFKSVGIFLGIFSGSFMTGAVTGVVTALISFLQNA	301	
Sbjct:	158	A+VL + +++ G A +A F + FL +F G +G V G + +LI	207	
NOV71:	302	NVTKFTKLHCFFPLETALFFLMSWSTFLLAECGFTGVVAVLFCGITQAHYTYNNLSVES	361	
Sbjct:	208	T+FT L+E L L+++ +LLAE G +C++AV G+ ++Y N+S +S	264	
NOV71:	362	RSRTKQLFEVLHFLAENFIFSYMGLALFTFQKHVFSPIFIIGAFVAIFLGRAAHYIPLSF	421	
Sbjct:	265	R+ K ++VL FL E IF +GL+L H ++ I+ A V I L RA ++ L+	324	
NOV71:	422	FLNLGRRHKIGWNFQHHMMFSGLRGAMAFALAI RDTASY--ARQMMFTTTLLIVFFT VWI	479	
Sbjct:	325	LN RR KI + Q ++ + GLRGA+A ALA+ + AR ++ TT +++V TV +	384	
NOV71:	480	IGGGTTPMLSWLNIR	494	
Sbjct:	385	G P++ L ++	399	

Na⁺/H⁺ exchangers are integral membrane ion transporter proteins that exchange extracellular Na⁺ for intracellular H⁺ with a stoichiometry of one for one. They have multiple cellular functions, including maintenance of intracellular pH, cell volume control, and reabsorption of sodium across renal, intestinal, and other epithelia. Multiple Na⁺/H⁺ exchanger isoforms (NHE1-NHE6) exist, exhibiting considerable differences in their

membrane localization, biochemical and pharmacologic properties, and responsiveness to various stimuli. For example, NHE1, the most predominant isoform expressed in heart, contributes significantly to myocardial intracellular pH. Hyperactivation of NHE1 during episodes of cardiac ischemia and reperfusion has been shown to disrupt the intracellular ion balance, leading to cardiac dysfunction and damage in several animal models, which can be prevented by pharmacological antagonists of NHEH. Increased activity of sodium/hydrogen exchange also provides a potentially important mechanism for the development of hypertension.

In 1998, Numata et al. identified the gene encoding a novel isoform of sodium/hydrogen exchanger that they called NHE6 or SLC9A6 (3). The NHE6 protein has similar topology to the other NHEs in that it has 12 putative membrane-spanning segments within the N-terminal region and a hydrophilic C terminus. However, NHE6 also has a putative mitochondrial inner membrane-targeting signal at its N terminus. The NOV71 protein is homologous to the NHE6 protein, except that it is predicted to localize to the plasma membrane instead of the mitochondrial inner membrane. The NHE6-like gene maps to human chromosome Xp11. Based on its expression pattern, NOV71 may play a role in renal or metabolic diseases and immune function through its sodium/hydrogen exchange activity at the plasma membrane.

NOV71 is predicted to be expressed in at least the following tissues: kidney, tonsils, germinal B cells, uterus, pituitary gland, brain, skeletal muscle, heart, lung, liver, pancreas, small intestine, colon, kidney, spleen, thymus, peripheral blood leukocytes, testis, ovary, placenta and prostate. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV71 is provided in Example 2.

The NOV71 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, cancer, trauma, bacterial and viral infections, in vitro and in vivo regeneration, endometriosis, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, diabetes, autoimmune disease, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalcaemia, Lesch-Nyhan syndrome, endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders, tonsillitis as well as other diseases, disorders and conditions.

The NOV71 nucleic acids encoding the sodium/hydrogen exchanger NHE6-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. The novel nucleic acid of the invention encoding a Sodium/Hydrogen Exchanger 6-like protein includes the nucleic acid whose sequence is provided in Table 71A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 71A while still encoding a protein that maintains its Sodium/Hydrogen Exchanger 6-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 71A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 34% of the bases may be so changed.

The novel protein of the invention includes the Sodium/Hydrogen Exchanger 6-like protein whose sequence is provided in Table 71B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 71B while still encoding a protein that maintains its Sodium/Hydrogen Exchanger 6-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 25% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV72

The disclosed NOV72 (alternatively referred to herein as CG56910-01) includes the 1094 nucleotide sequence (SEQ ID NO:247) shown in Table 72A. A NOV72 ORF begins

Kozak consensus ATG initiation codon at nucleotides 14-16 and ends with a stop codon at nucleotides 1082-1084. The disclosed NOV72 maps to human chromosome 6.

Table 72A. NOV72 Nucleotide Sequence (SEQ ID NO:247)

```
ACTTCTATAAGACATGGATAGATGCAAACATGTAGGGCGGTTAGCCCCAGGCGTCACGGG
CCTGCGCAACCTGGGCAACACCTGCTACATGAACTCCATCCTCCAGGTGCTCAGCCACCT
CCAGAAGTTCCGAGAATGTTTCTCAACCTTGACCCCTTCCAAAACGGAACATCTGAGTTC
AAAGCACATTCCCTCTGCCGTGAAGTGCACACCCTCTCCGAGTCATGTGGTCCGGGAA
GTGGGCCCTAGTGTGCGCCCTTCGCCATGCTCCACTCAGTGTGGAGCCTGATCCCTGCCTT
CCGCGGCTACGACCAACAGGACGCGCAGGAATTTCTCTGCGAGCTGCTGCACAGGTGCA
GCAGGAACTCGAGTCTGAGGGCACACGCGGATCCTCATCCCCTTCTCCAGAGGAA
GCTACCAACAGGTCTTAAAGGTGGTGAATACCATATTTTCATGGGCAGCTGCTCAGTCA
GGTATGTGTGGTCACATGTATATCATGCAATTACAAATCCAATACCATTGAGCCCTTTTG
GGACCTATCCCTGGAATTCCTGAAACGCTATCACTGCATAGAAAAGGGGTTTGTCCCTTT
GAATCAAACAGAGTGCCTTGCTCACTGAGATGCTGGCCAAATTCACAGAGACAGAGGCCCT
GGAAGGGAGAATCTACGCTTGTGACCAGTGTAAACAGTGAAGTGTGTGAAGCAGTTAAT
GATCTACAGACTACCTCAGGTTCTCCGGCTGCACCTTAAAAGATTCCATCGAGAGAAGAT
TGGGGTCCATGTCGTCTTTGACCAGGTATTAACCATGGAACCTTACTGCTGCAGGGACAT
GCTCTCCTCTCTTGACAAAGAGACCTTTGCCTATGATCTCTCCGAGTGGTCAATGCAATCA
CGGAAAGGGTTTGGCTCAGGACACTACACAGCCTATTGCTACAACACAGAGGGAGGTTT
TTGGGTCCACTGCAATGACTCAAAGCTGAATGTATGCAGTGTGCGAGGAAGTGTGCAAAAC
CCAGGCCCTACATCCTTTTACGCGCTGACTGAGATGGCGCTGAGTGAATGTGAAGGTG
CTAAGACCCAGTCT
```

5 A NOV72 polypeptide (SEQ ID NO:248) encoded by SEQ ID NO:247 is 356 amino acids in length and is presented using the one-letter amino acid code in Table 72B. The Psort profile for NOV72 predicts that this sequence has no signal peptide and is likely to be localized in the cytoplasm with a certainty of 0.5050. In alternative embodiments, a NOV72 polypeptide is located to lysosomes with a certainty of 0.1000, to the mitochondrial matrix space with a certainty of 0.1000, or to peroxisomal microbodies with a certainty of 0.3547.

Table 72B. NOV72 Polypeptide Sequence (SEQ ID NO:248)

```
MDRCKHVGR LAPGV TGLRNLGNTCYMNSILQVLSHLQKFRECFLNLDPSKTEHLSKHS
LCRELHTLFRVMWSGKVALVSPFAMLSVWSLIPAFRGYDQQAQEFLLCELLHKVQEELE
SEGTTTRRILIPFSQRKLTKQVLKVNTIFHGQLLSQVCVVTCISCNYSNTIEPFWDLSL
EPPERHCEIEKGFVPLNQTECLLTEM LAKFTETEAL EGR IYACDQCNSECCVKQLMIYRL
PQVLRRLHLKRFRHREKIGVHVVFQVLTMEPYCCRDMLSSLDKETFA YDLSAVVMH HGKGF
GSGHYTAYCYNTEGGFWVHCNDSKLNVCVVEEVCKTQAYILFYALTEMALSECGR
```

15 A BLAST analysis of NOV72 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV72 had high homology to other proteins as shown in Table 72C.

Table 72C. BLASTX results from PatP database for NOV72

Smallest

Sequences producing High-scoring Segment Pairs:			High Score	Sum Probability P(N)
patp:AAB92670	Human protein sequence		1133	2.0e-121
patp:AAB42259	Human ORFX ORF2023 polypeptide sequence		497	2.7e-47
patp:AAY13115	Human secreted protein encoded by 5' EST		492	9.1e-47
patp:AAM79194	Human protein		487	3.1e-46
patp:AAB74672	Human protease and protease inhibitor		464	8.4e-44

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 215 of 342 bases (62%) identical to a gb:GENBANK-ID:AF073344|acc:AF073344.1 mRNA from *Homo sapiens* (ubiquitin-specific protease 3 (USP3) mRNA). The full amino acid sequence of the protein of the invention was found to have 124 of 345 amino acid residues (35%) identical to, and 183 of 345 amino acid residues (53%) similar to, the 353 amino acid residue ptnr:SWISSNEW-ACC:O88623 protein from *Mus musculus* (Mouse) (UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 2 (EC 3.1.2.15) (UBIQUITIN THIOLESTERASE 2) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 2) (DEUBIQUITINATING ENZYME 2) (41 KDA UBIQUITIN-SPECIFIC PROTEASE)). NOV72 also has homology to the other proteins shown in the BLASTP data in Table 72D.

Table 72D. NOV72 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 15208127 d bj BAB63088.1 (AB070143)	hypothetical protein [Macaca fascicularis]	550	282/317 (88)	287/317 (89)	e-157
gi 16174134 r ef XP_057397. 1 (XM_057397)	similar to Unknown (protein for MGC:20741) (<i>H. sapiens</i>)	640	255/289 (88)	260/289 (89)	e-140
gi 14149817 r ef NP_115523. 1 (NM_032147)	hypothetical protein DKFZp434D0127 [<i>Homo sapiens</i>]	712	221/307 (71)	247/307 (79)	e-119
gi 15451368 d bj BAB64488.1 (AB071094)	hypothetical protein [Macaca fascicularis]	585	205/239 (85)	210/239 (87)	e-108
gi 16041104 d bj BAB69719.1 (AB072750)	hypothetical protein [Macaca fascicularis]	497	111/213 (52)	123/213 (57)	2e-45

This BLASTP data is displayed graphically in the ClustalW in Table 72E. A multiple sequence alignment is given, with the NOV72 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 72D.

Table 72E. ClustalW Alignment of NOV72

NOV72	(SEQ ID NO:248)
gi 15208127	(SEQ ID NO:659)
gi 16174134	(SEQ ID NO:660)
gi 14149817	(SEQ ID NO:661)
gi 15451368	(SEQ ID NO:662)
gi 16041104	(SEQ ID NO:663)

	10	20	30	40	50
NOV72	----	----	----	----	----
gi 15208127	---	MDRCKHVGRLRLA	QDHSILNPQKWC	CLECATTESVWAC	LKCSHVACG
gi 16174134	MLAMDTCKHVG	QLAQDHSSLN	PQKWHCVDCN	TTESTIWACL	SCSHVACG
gi 14149817	---	MDRCKHVGRLRLA	QDHSILNPQKWC	CLECATTESVWAC	LKCSHVACG
gi 15451368	---	MDRCKHVGRLRLA	QDHSILNPQKWC	CLECATTESVWAC	LKCSHVACG
gi 16041104	MLTMDKCKHVG	QLAQDHSSLN	PQKWHCVDCN	TTESTIWACL	SCSHVACG

	60	70	80	90	100
NOV72	----	----	----	----	----
gi 15208127	RYIEDHALKHFE	ETGHPLAMEVR	DLVVF	CYLCKDYVL	NDNPEGDLKLLRS
gi 16174134	RYIEBHALKHF	QESSHPVALE	VNEMVVF	CYLCKDYVL	NDNATGDLKLLRR
gi 14149817	RYIEDHALKHFE	ETGHPLAMEVR	DLVVF	CYLCKDYVL	NDNPEGDLKLLRS
gi 15451368	RYIEBHALKHF	QESSHPVALE	VNEMVVF	CYLCKDYVL	NDNATGDLKLLRS
gi 16041104	RYIEBHALKHF	QESSHPVALE	VNEMVVF	CYLCKDYVL	NDNATGDLKLLRS

	110	120	130	140	150
NOV72	----	----	----	----	----
gi 15208127	SLLAVRGQKQ	DTPVRRGR	TLRSMASGED	----	MLTAL
gi 16174134	TLTALKSONYH	CTTRSGR	FLRSMGTG	DDSYFLHDGA	QSLLOSEDOLYTAL
gi 14149817	SLLAVRGQKQ	DTLVRGR	TLRSTASGED	----	VVPPORAPQGPOMLTAL
gi 15451368	SLLAVRGQKQ	DTLVRGR	TLRSTASGED	----	VVPPORAPQGPOMLTAL
gi 16041104	TLTALKSONYH	CMTRSGM	VLRSMTG	SDSYFLYDGA	QSLLOEDOMYTAL

	160	170	180	190	200
NOV72	----	----	----	----	----
gi 15208127	WYRRORLLART	LRLWF	EKSSRCQAKLE	ORROBEAL	ERKKEEARRRRREV
gi 16174134	WYRRORLLART	LRLWF	EKSSRCQAKLE	ORROBEAL	ERKKEEARRRRREV
gi 14149817	WHRRRILMGK	IFRTWFEQ	SPIGRKKQ	EEPFQEKLV	VVKREVKRRRQ--ELE
gi 15451368	WYRRORLLART	LRLWF	EKSSRCQAKLE	ORROBEAL	ERKKEEARRRRREV
gi 16041104	WHRRRILMGK	IFRTWFEQ	SPIGRKKQ	EEGFQEKLV	VVKREVKRRRQ--ELE

	210	220	230	240	250
NOV72	----	----	----	----	----
gi 15208127	RRILLEELAS	APPRKSARLL	----	LHTESDAGP	AASRPPTTLPT
gi 16174134	RRILLEELAS	TPPRKSARLL	----	LHTERDAGP	AASRPAAALPT
gi 14149817	YQVKAEL	ESMPPRKSL	RLOGLAQST	IIEIVSVQV	PAQTPASPAKDKVLST
gi 15451368	RRILLEELAS	APPRKSARLL	----	LHTESDAGP	AASRPPTTLPT
gi 16041104	YQVTAEL	ESMPPRKSL	RLOGLAQST	IIEIVSVQV	PAQMPASTAKDKVLST

	260	270	280	290	300
NOV72	----	----	----	----	----
gi 15208127	SRRAPAA	TIK---LR	ROPAP	APGVTGL	RNLGNTCYMNSILQVLSHLQKFR
gi 16174134	SRRVPA	ATI---LR	ROPAP	APGVTGL	RNLGNTCYMNSILQVLSHLQKFR
gi 14149817	SENEIS	QKVS	DSSV	KRRPIV	TPGVTGLRNLGNTCYMNSVLOVLSHLIIFR
gi 15451368	SRRAPAA	TIK---LR	ROPAP	APGVTGL	RNLGNTCYMNSILQVLSHLQKFR
gi 16041104	SEDER	SQK	INDSS	VKRRPIV	TPGVTGLRNLGNTCYMNSVLOVLSHLIIFR

	310	320	330	340	350
NOV72	----	----	----	----	----
gi 15208127	ECFLNLD	PSKTEHL	----	----	----

gi	15208127	ECFLNLDPSKTEHLFPKATNGKTQLSGKPTNSSATELSLRSDRAEACERE
gi	16174134	ECFLNLDPSKTEHLFPKATNGKTQLSGKPTNSSATELSLRNDRAEACERE
gi	14149817	QCFLKLDLNQWLAMTASEKTRS-CKHPPVTDIVVYQMNCEQEKDTGFVCS
gi	15451368	ECFLNLDPSKTEHLFPKATNGKTQLSGKPTNSSATELSLRSDRAEACERE
gi	16041104	QCFLKLDLNQWLAVTASEKKRSFSKHPPVTDIVVYQMNCEQEKDTGFVRS
		360 370 380 390 400
NOV72	SSKHISLCRELHTLFRVMWSGKW
gi	15208127	GFC----WNGGASISRSELIQNEPSSKHISLCRELHTLFRVMWSGKW
gi	16174134	GFC----WNGRASISRSELIQNEPSSKHISLCRELHTLFRVMWSGKW
gi	14149817	RQSSLSSGLSGGASKGRKMELIQPEPTISOYISLCHELHTLFRVMWSGKW
gi	15451368	GFC----WNGGASISRSELIQNEPSSKHISLCRELHTLFRVMWSGKW
gi	16041104	RQSSLSSGLSGGASKSRKMELIQPREPTISOYISLCHELHTLFRVMWSGKW
		410 420 430 440 450
NOV72		ALVSPFAMLHSVWSLIPAFRGYDQDDAQEFLCELLHKVQOELESEGTTTR
gi	15208127	ALVSPFAMLHSVWSLIPAFRGYDQDDAQEFLCELLHKVQOELESEGTTTR
gi	16174134	ALVSPFAMLHSVWSLIPAFRGYDQDDAQEFLCELLHKVQOELESEGTTTR
gi	14149817	ALVSPFAMLHSVWRLIPAFRGYAQQDAQEFLCELLDKIQRELETTTGTSLP
gi	15451368	ALVSPFAMLHSVWSLIPAFRGYDQDDAQEFLCELLHKVQOELESEGTTTR
gi	16041104	ALVSPFAMLHSVWRLIPAFRGYAQQDAQEFLCELLDKIQHELETTTGTSLP
		460 470 480 490 500
NOV72		ILIPFSQRKLTKQVLKVNTIFHGQLLSQVCVTCISCNYKSNTIEPFWD
gi	15208127	ILIPFSQRKLTKQVLKVNTIFHGQLLSQ---VTCISCNYKSNTIEPFWD
gi	16174134	ILIPFSQRKLTKQVLKVNTIFHGQLLSQ---VTCISCNYKSNTIEPFWD
gi	14149817	ALIPFSQRKLTKQVLNVVNNIFHGQLLSQVTCISCNYKSNTIEPFWD
gi	15451368	ILIPFSQRKLTKQVLKVNTIFHGQLLSQ---VTCISCNYKSNTIEPFWD
gi	16041104	ALIPFSQRKLTKQVLNVVNNIFHGQLLSQ-----SVEG-----
		510 520 530 540 550
NOV72		LSLEFPERYHCIEKGFVPLNTECLLTEMLAKFTETEALG-----RIY
gi	15208127	LSLEFPERYHCIEKGFVPLNTECLLTEMLAKFTETEALG-----GIY
gi	16174134	LSLEFPERYHCIEKGFVPLNTECLLTEMLAKFTETEALG-----RIY
gi	14149817	LSLEFPERYQCSGK---DIASQPCLVTEMLAKFTETEALGCKIYVCDQCN
gi	15451368	LSLEFPERYHCIEKGFVPLNTECLLTEMLAKFTETEALG-----RIY
gi	16041104	---FPP-----
		560 570 580 590 600
NOV72		ACDQCNSC-----CVKQLMIYRLPQVLRHLHLKRF-----HR
gi	15208127	ACDQCNSKRRKSNPKPLVLSEARKQLMIYRLPQVLRHLHLKRFWSGRNHR
gi	16174134	ACDQCNSKRRKSNPKPLVLSEARKQLMIYRLPQVLRHLHLKRFWSGRNHR
gi	14149817	SKRRRFSSKP-----VLTEAQQLMICHLPQVLRHLHLKRFWSGRNHR
gi	15451368	ACDQCNSKRRKSNPKPLVLSEARKQLMIYRLPQVLRHLHLKRFWSGRNHR
gi	16041104	--NOLYSQK-----P---RNNL
		610 620 630 640 650
NOV72		EKIGVHVVDQVLTMEPYCCRDMLSSLDKETAYDLSAVVMHHGKGFSG
gi	15208127	EKIGVHVVDQVLTMEPYCCRDMLSSLDKETAYDLSAVVMHHGKGFSG
gi	16174134	EKIGVHVVDQVLTMEPYCCRDMLSSLDKETAYDLSAVVMHHGKGFSG
gi	14149817	EKIGVHVGFEEILNMEPYCCRETLKSLRPECFTYDLSAVVMHHGKGFSG
gi	15451368	EKIGVHVVDQVFGST-----AMTQS-----
gi	16041104	-----
		660 670 680 690 700
NOV72		HYTAYCYNTEGGFWVHCNDSKLNVCSSVEEVCKTQAYILFY-----
gi	15208127	HYTAYCYNTEGGFWVHCNDSKLNVCSSVEEVCKTQAYILFYTQRTVQGNAR
gi	16174134	HYTAYCYNTEGGACALLCGVGDTERG-----
gi	14149817	HYTAYCYNSEGGFWVHCNDSKLSMCTMDEVCKAQAYILFYTQRTVENTGHS
gi	15451368	-----
gi	16041104	-----

```

                                710          720
      . . . . | . . . . | . . . . | . . . . | . . . . | . .
NOV72      -----ALTEMALSECGR-----
gi|15208127| ISETHLQAQVQSSNNDEGRPQTFS---
gi|16174134| -----
gi|14149817| KLLPPELLLSQHNPEDADTSSNEILS
gi|15451368| -----
gi|16041104| -----

```

Table 72F lists the domain description from DOMAIN analysis results against NOV72. This indicates that the NOV72 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 72F. Domain Analysis of NOV72			
<u>gnl</u> <u>Pfam</u> <u>pfam00443</u> , UCH-2, Ubiquitin carboxyl-terminal hydrolase family 2. SEQ ID NO:866 CD-Length = 68 residues, 91.2% aligned Score = 59.7 bits (143), Expect = 3e-10			
NOV72:	287	YDLSAVVMHHGKGFGSGHYTAYCYNTEGGFWVHCNDSKLVCSVEEVCK-----TQAY	339
		Y+L AVV+H G GHY AY G W +D K++ + EEV + + AY	
Sbjct:	5	YELYAVVVHSG-SLSGGHYIAYVKKENDG-WYKFDDDKVSRVTEEEVLEFGSGGETSSAY	62
NOV72:	340	ILFY	343
		ILFY	
Sbjct:	63	ILFY	66

Ubiquitin is a highly conserved 76-amino acid protein involved in the regulation of intracellular protein breakdown, cell cycle regulation, and stress response. Ubiquitin is released from degraded proteins by disassembly of the polyubiquitin chains, which is mediated by ubiquitin-specific proteases (USPs). The ubiquitin-specific proteases are a family of largely dissimilar enzymes with two major conserved sequence regions, containing either a conserved cysteine residue or two conserved histidine residues, respectively. The murine Unp oncoprotein and its human homologue, Unph, both contain regions similar to the conserved Cys and His boxes common to all the Ubps.

Unp and Unph have been shown to be active deubiquitinating enzymes, able to cleave ubiquitin from both natural and engineered linear ubiquitin-protein fusions, including the polyubiquitin precursor. Mutation of the conserved Unp Cys and His residues abolishes this activity, and identifies the likely His residue in the catalytic triad. Unp is tumorigenic when overexpressed in mice, leading to the suggestion that Unp may play a role in the regulation of ubiquitin-dependent protein degradation. It was demonstrated that the high-level expression of Unp in yeast does not disrupt the degradation of the N-end rule substrate Tyr-beta-

galactosidase (betagal), the non-N-end rule substrate ubiquitin-Pro-betagal, or the degradation of abnormal, canavanine-containing proteins.

Data suggests that Unp is not a general modulator of ubiquitin-dependent proteolysis. However, Unp may have a role in the regulation of the degradation of a specific, as yet undescribed, substrate(s). The novel human Ubiquitin-Specific Protease-like Proteins of the invention contains two Ubiquitin carboxyl-terminal hydrolase domains. Therefore, it is anticipated that NOV72 has a role in regulation of specific ubiquitins and could be a potentially important target for drugs. Such drugs may have important therapeutic applications, such as treating numerous tumors.

NOV72 is predicted to be expressed in at least the following tissues: bladder and cervix. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV72 is provided in Example 2.

The NOV72 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: cystitis, incontinence, fertility, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, as well as other diseases, disorders and conditions. The NOV72 nucleic acid encoding the ubiquitin specific protease-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a ubiquitin-specific protease-like protein includes the nucleic acid whose sequence is provided in Table 72A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 72A while still encoding a protein that maintains its ubiquitin-specific protease-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 72A, including nucleic acid fragments that are complementary to any of

the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 38% of the bases may be so changed.

The novel protein of the invention includes the ubiquitin-specific protease-like protein whose sequence is provided in Table 72B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 72B while still encoding a protein that maintains its ubiquitin-specific protease-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 65% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV73

The disclosed NOV73 (alternatively referred to herein as CG56822-01) includes the 967 nucleotide sequence (SEQ ID NO:249) shown in Table 73A. A NOV73 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 26-28 and ends with a stop codon at nucleotides 935-937. The disclosed NOV73 maps to human chromosome 2.

Table 73A. NOV73 Nucleotide Sequence (SEQ ID NO:249)

```
GCAGGTTCTTTTACAGGGAGCCACCATGGCTGATAAATCCAAATTTATTGAATACATTGA
CGAAGCTTTAGAAAAATCAAAGAACTGCACCTCTCTCATTATTTTTCACCTATCAGGG
GATTCCTTACCCCATCACCATGTGCACCTCAGAACTTTCCAAGCGCTGGACACCTTCGA
AGCCAGACATGATGACATCGTGCTAGCATCTTATCCAAAGTGCAGTTCAAACCTGGATTCT
CCACATTGTCAAGTAATTAATATATGCTGTTTCTAAAAAAGTATAAATATCCAGAATT
CCCAGTTCCTTGAATGTGGGGATTCAGAAAAATATCAGAGAATGAAAGGCTTCCATCACC
AAGGATTTTGGCAACTCACCTCCACTATGACAAATTACCTGGGTCTATCTTCGAGAATAA
AGCCAAGATATTGGTGATATTTGAAACCCTAAAGATACAGCAGTATCTTTTTTGCATTT
CCACAACGATGTCCCGGATATTCAGCTATGGCTCTTGGGATGAATCTTCAGACAGTT
CATGGTGTTTTAGTTTCTTGGGGAAGGTATTTGATTTTGCAATCAATTGGAACAAACA
TCTTGATGGCGACAATGTTAAGTTCATATTATATGAAGACCTGAAAGAGGTGAGATTATT
AGGAATAAACAGATTGCTGAGTTCTTGGGATTCTTCTAACTGGGGAGCAAATTCAAAC
```

TATCTCAGTCCAGAGCACCTTCCAAGCCATGCGTGCGAAGTCTCAGGACACACACGGTGC
TGTCGGCCCATTCCTTTTCCGCAAAGGTAAAGTCGCAGATTGGAATAATTGTTTCAGTGA
AATTCAGAACCCAGGAAATGGATGAAAAATTCAAAGAGTGCTTAGCAGGCACCTCCCTCGG
AGCAAAGTTGAAGTATGAATCATATTGCCAGGGTTGATTCCAGTCAATTCAGCAGGCCTA
GATTTAT

A NOV73 polypeptide (SEQ ID NO:250) encoded by SEQ ID NO:249 is 303 amino acids in length and is presented using the one-letter amino acid code in Table 73B. The Psort profile for NOV73 predicts that this sequence has no signal peptide and is likely to be localized to peroxisomal microbodies with a certainty of 0.7480. In alternative embodiments, a NOV73 polypeptide is located to lysosomes with a certainty of 0.1000, or to the mitochondrial matrix space with a certainty of 0.1000.

Table 73B. NOV73 Polypeptide Sequence (SEQ ID NO:250)

MADKSKFIEYIDEALEKSKETALSHLFFTYQGI PYPITMCTSETFQALDTFEARHDDIVL
ASYPKCGSNWILHIVSELIYAVSKKKYKYPEFPVLECGDSEKYQRMKGFPSPRILATHLH
YDKLPGSIFENKAKILVIFRNPKDTAVSFLHFHNDVDPISYGSWDEFFRQFMVFLVSWG
RYFDFAINWNKHLGDGNVKFIFYEDLKEVRLGKQIAEFLGFFLTGEQIQITISVQSTFQ
AMRAKSQDTHGAVGPFLFRKGKVADWKNLFSEIQNQEMDEKFKCELAGTSLGAKLKYESY
CQG

A BLAST analysis of NOV73 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV73 had high homology to other proteins as shown in Table 73C.

Table 73C. BLASTX results from PatP database for NOV73

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAU07758 Human novel transferase protein,	1562	3.7e-160
patp:AAU07760 Human novel transferase protein,	1366	2.2e-139
patp:AAU07765 Human novel transferase protein,	1013	5.6e-102
patp:AAU07763 Human novel transferase protein,	930	3.5e-93
patp:AAU07761 Human novel transferase protein,	734	2.1e-72

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 427 of 643 bases (66%) identical to a gb:GENBANK-ID:AF033189|acc:AF033189.1 mRNA from Gallus gallus (Gallus gallus sulfotransferase mRNA). The full amino acid sequence of the protein of the invention was found to have 151 of 307 amino acid residues (49%) identical to, and 212 of 307 amino acid residues (69%) similar to, the 312 amino acid residue ptnr:SPTREMBL-ACC:O57338 protein from Gallus

gallus (Chicken) (SULFOTRANSFERASE). NOV73 also has homology to the other proteins shown in the BLASTP data in Table 73D.

Table 73D. NOV73 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17447308 ref XP_065865.1 (XM_065865)	similar to sulfotransferase (H. sapiens) [Homo sapiens]	303	294/303 (97)	295/303 (97)	e-168
gi 2687360 gb AAB88818.1 (AF033189)	sulfotransferase [Gallus gallus]	312	151/307 (49)	212/307 (68)	4e-84
gi 12229955 sp Q9WUW8 STK1_RAT	SULFOTRANSFERASE K1 (RSULT1C2)	296	94/294 (31)	146/294 (48)	3e-35
gi 18079235 ref NP_081211.1 (NM_026935)	sulfotransferase family, cytosolic, 1C, member 1 [Mus musculus]	296	92/292 (31)	147/292 (49)	5e-35
gi 11262122 pir JC7283	hydroxyarylamine sulfotransferase (EC 2.8.2) 2A - rat	296	93/294 (31)	149/294 (50)	2e-34

- 5 This BLASTP data is displayed graphically in the ClustalW in Table 73E. A multiple sequence alignment is given, with the NOV73 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 73D.

Table 73E. ClustalW Alignment of NOV73	
NOV73	(SEQ ID NO:250)
gi 17447308	(SEQ ID NO:664)
gi 2687360	(SEQ ID NO:665)
gi 12229955	(SEQ ID NO:666)
gi 18079235	(SEQ ID NO:667)
gi 11262122	(SEQ ID NO:668)
<div> <div>1020304050</div> <div> NOV73MAD-KSKFIEYIDEALEKSKETALSHLFFTYQGIPYELITMCTSETFOALD gi 17447308 MAD-KSKFIEYIDEALEKSKETALSHLFFTYQGIPYELITMCTSETFOALD gi 2687360 MEKSRKKFIDVTDKAIIVIGNAMDRDEILFSYKGVLYEVALCSPVFRAME gi 12229955 -----MALAPELSRQTKLKEVAG-----IPLQAPTVDNWSQIQ gi 18079235 -----MALTPELSRQTKLKEVAG-----IPLQAPTVDNWSQIQ gi 11262122 -----MALAPELSRQTKLKEVAG-----IPLRDSTVDNWSQIQ </div> </div>	
<div> <div>60708090100</div> <div> NOV73TFEARHDDIVLASYPKCGSNWILHIVSELIYAVSKKKYK-----YPE gi 17447308 TFEARHDDIVLASYPKCGSNWILHIVSELIYAVSKKKYK-----YPE gi 2687360 SFEARSDDVILAGYPKSGTNWVGCILSDLVATFEKERLEEKSVNDEELEE gi 12229955 TFKAKPDDLILCIYPKSGTTWIOEIVDMTEQNGDVEKCOR-TIIQHRHPF gi 18079235 TFEAKPDDLILCIYPKSGTTWIOEIVDMTEQNGDVEKCOR-TIIQHRHPF gi 11262122 TFKAKPDDLILCIYPKSGTTWIOEIVNMTEQNGDVEKCOR-TIIQHRHPF </div> </div>	
<div> <div>110120130140150</div> <div> </div> </div>	

NOV73	FPVLECGDSEKYQRMKGFPSPRILATHLHYDKLPGSIFENKAKILVIFRN
gi 17447308	FPVLECGDSEKYQRMKGFPSPRILATHLHYDKLPGSIFENKAKILVIFRN
gi 2687360	FPYLEIGDTEKYERMKKLPSRRVILTHLSPEKLPKSIENKAKILLIRN
gi 12229955	IEWARPPQPSGVDDKANAMPAPRILRTHLPTQLLPSPFWTNNCKFLYVARN
gi 18079235	IEWARPPQPSGVDDKANAMPAPRILRTHLPTQLLPSPFWTNNCKFLYVARN
gi 11262122	IEWARPPQPSGVDDKANAMPAPRILRTHLPIQLLPSPFWTNNCKFLYVARN
	160 170 180 190 200
NOV73	PKDTAVSFLHFHNDVPDIPSYGSWDEFFRQFMVFLVSWGRYFDFAINWNK
gi 17447308	PKDTAVSFLHFHNDVPDIPSYGSWDEFFRQFMKQVSWGRYFDFAINWNK
gi 2687360	PKDIATSFHHSNRWSALPSYETWDDFFIAFMTEKMPWGSYFNLYSEWNK
gi 12229955	AKDCMVSYHFRMSQVLEDPGTWNEYFETFLNGKVSWSGSFDFHVKGWWE
gi 18079235	AKDCMVSYHFRMSQVLEDPGTWDEYFETFLNGKVSWSGSFDFHVKGWWE
gi 11262122	AKDCMVSYHFRMCQVLENPCTWNEYFETFLNGKVSWSGSFDFHVKGWWE
	210 220 230 240 250
NOV73	HLDGDNVVKFIFYEDLKEVRLLGKQIAEFLGFFLTGEQIQTHSVQSTFQA
gi 17447308	HLDGDNVVKFIFYEDLKENLAAGIKQIAEFLGFFLTGEQIQTHSVQSTFQA
gi 2687360	YAADENVMTITYEELKENQTLGVKNLASFFGISTLGEELRSVIERSSFQS
gi 12229955	IRDRYQILFLFYEDVKRDPKREIQKVMQFMGKNLDEEVVDKIVLETSEK
gi 18079235	IRDKYQILFLFYEDMKRNPKEIQKVMQFMGKNLDEDVVDKIVLETSEK
gi 11262122	IRDRYQILFLFYEDMKRDPKREIQKVMQFMGKNLDEEVVDKIVLETSEK
	260 270 280 290 300
NOV73	MRAKSQDTHGAVG-----P---FLFRKGKVADWKNLFSEIQNQEMDEKFK
gi 17447308	MRAKSQDTHGAVG-----P---FLFRKGEVGDWKNLFSEIQNQEMDEKFK
gi 2687360	MKENSILKTHGALG-----S---MLFRKGCVSQWKNLFNEEQNEKMDKVEE
gi 12229955	MKENPMTNRSTYPKSVLDQSSISPFMRKGTIVGDWKNHFTVAQNDRFDEEYK
gi 18079235	MKENPMTNRSTAPKSILDQSSISPFMRKGTIVGDWKNHFTVAQNERFDEEYK
gi 11262122	MKDNPLTNFSTLPKTIMDQSSISPFMRKGTIVGDWKNHFTVAQNERFDEEYK
	310 320
NOV73	ECLAGTSLGAKLKYESYCOG
gi 17447308	ECLAGTSLGAKLKYESYCOG
gi 2687360	ERIARTKLGTGLKYEVYCKA
gi 12229955	QKMGTSLSNFCMEL-----
gi 18079235	QKMGRTSLNFSMEL-----
gi 11262122	QKMDGTSLSNFCMEL-----

Table 73F lists the domain description from DOMAIN analysis results against NOV73. This indicates that the NOV73 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 73F. Domain Analysis of NOV73				
gnl Pfam	pfam00685,	Sulfotransfer,	Sulfotransferase protein	SEQ ID NO:867
CD-Length = 269 residues, 93.3% aligned				
Score = 176 bits (446), Expect = 2e-45				
NOV73:	49	DTFEARHDDIVLASYPKCGSNWLHIVSELIYAVSKKKYKYPE-----FPVLECGDSEK-	102	
Sbjct:	18	+ F+AR DD+++A YPK G+ W+ I+S L V + + P L E E	76	
NOV73:	103	YQRMKGFPSPRILATHLHYDKLPGSIFENKAKILVIFRNPKDTAVSFLHFHNDVPDIPS	161	
Sbjct:	77	Y+ +K PS PR++ THL + LP S +KAKI+ + RNP KD AVS+ HF D+P+ YETLKMPSPSPRLIKTHLPLELLPKSFLSSKAKIIYVLRNPKDVAVSYYHFSRSHKDLPA	136	
NOV73:	162	Y-GSWDEFFRQFMVFLVSWGRYFDFAINWNKHLDGDNVVKFIFYEDLKEVRLLGKQIAEF	220	

		G+++EF	F+	V +G YFD	+ W +	V F+ YEDLKE	IK+IAEF	
Sbjct:	137	DPGTFEEFLEAF	LNGKVL	YGSYFDHVL	GWELRPEP	QVLF	LDYEDLKEDPAGEIKKIAEF	196
NOV73:	221	LGFFLTGEQIQ	TISVQSTFQ	AMRA-----	KSQDTHGAV	GPFLFRK	GKVADWKNLFS	271
		LG L+ E++	+ S+F M+		+ G PF	RKG V	DWKN F+	
Sbjct:	197	LGLPLSEEEL	DKLLDHSS	FFLMKNPL	SNYETLCL	GKSKGRK	SPF-MRKGLVGDWKNYFT	255
NOV73:	272	EIQNQEMDEK	FKKE	284				
		QN++ D+	KE					
Sbjct:	256	PEQNEKFDK	VIKE	268				

This family includes a range of sulfotransferase proteins including flavonyl 3-sulfotransferase, aryl sulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase and phenol-sulfating phenol sulfotransferase. These enzymes are responsible for the transfer of sulphate groups to specific compounds.

NOV73 is predicted to be expressed in at least the following tissues: epithelial, endothelial, muscle, and neuronal tissues. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV73 is provided in Example 2.

The novel nucleic acid of the invention encoding a sulfotransferase-like protein includes the nucleic acid whose sequence is provided in Table 73A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 73A while still encoding a protein that maintains its sulfotransferase-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 73A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 34% of the bases may be so changed.

The novel protein of the invention includes the Sulfotransferase-like protein whose sequence is provided in Table 73B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 73B while

still encoding a protein that maintains its Sulfotransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 51% of the amino acid residues may be so changed.

5 The NOV73 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: cystitis, incontinence, fertility, systemic lupus erythematosus, autoimmune disease, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, 10 Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration. as well as other diseases, disorders and conditions. The NOV73 nucleic acid encoding the sulfotransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid 15 or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section 20 below.

NOV74

25 The disclosed NOV74 (alternatively referred to herein as CG56775-01) includes the 732 nucleotide sequence (SEQ ID NO:251) shown in Table 74A. A NOV74 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 52-54 and ends with a stop codon at nucleotides 673-675. The disclosed NOV74 maps to human chromosome 15.

Table 74A. NOV74 Nucleotide Sequence (SEQ ID NO:251)

CAGACTCCCCCTTGCTGGCTCCTGCACAGAATGCCGGGCCCCACTGCTGCCATGACAGGCCCTTTCAAAC GCTCCATGGAAGATCTCAGTGACCTGCTCTCAGACAGCAGTGGCTGCTACAGCCTCCCAAGCCAGCCCTG CAATGAGGTCAACCCTGAAGATTTACATGGGCAACACATCTGTGGATCAGGATATCCCCAAGCTTCAGAAA CTAGGCAGTATTATGCCCTGAATACCACTGAGGGCAGGTCTTTCATGCACATAAACAATGCCAATTCT CCAAGGACTCTGGCATCACCTACCTGGGCATCAAGGCCAATGAAGTGCAGGAGTTCAACCTCAGCACCTA CTTTGAAAGGGCTACAGACTTCACTGACCAGGCCTTGGCTCAAAATGGCCAGGTGCTCGTCCAGTGCTGG GAAGGTTACAGCCATCTCCAGCTCGTTATCATGTACCTTATGATTGTCAGAAGTTGGACATCAAGTCATC TGAGTATCATGAGGCAGAACTGTGAGATCAGCCCCAATGATGGGTTCCTGGCTCAGCTTTGCCATCTCAA TGACAACTAGCCAAGGAGGGGAAGGTGAAACCCTGGGGTGCCCCCTACCACCTTTGCTCGAGAGGTTGAG TGGGAGAGGCCCTGGTTGAAGGTATCCTGTGACACTGTACCCTGATCCCAGCATCACGAGCCACTTGCCC TCAAGTCTGTCTCAACAAGTCTCTGGGCCACTT
--

A NOV74 polypeptide (SEQ ID NO:252) encoded by SEQ ID NO:251 is 214 amino acids in length and is presented using the one-letter amino acid code in Table 74B. The Psort profile for NOV74 predicts that this sequence has no signal peptide, and is likely to be localized at the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV74 polypeptide is located to peroxisomal microbodies with a certainty of 0.3625, or, to the mitochondrial matrix space with a certainty of 0.1000.

Table 74B. NOV74 Polypeptide Sequence (SEQ ID NO:252)

MPGPTAAMTGPFKRSMEDLSDLLSDSSGCYSLPSQPCNEVTLKIYMGNTSVDQDIPKLQK
LGSIHALNTTEGRSFMHINNANFSKDSGITYLGIKANEVQEFNLSTYFERATDFTDQALA
QNGQVLVQCWEGYSHLQLVIMYLMIVRSWTSSHLSIMRQNCIEISPNDGFLAQLCHLNDKL
AKEGKVKPWGAPTTFAREVQWERPWLKVSCTVP

A BLAST analysis of NOV74 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV74 had high homology to other proteins as shown in Table 74C.

Table 74C. BLASTX results from PatP database for NOV74

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAR56968 Human phosphatase VHR - <i>Homo sapiens</i> , 185 aa.	664	5.4e-65
patp:AAW35330 Human cdc25B vaccinia H1 related phosphatase	664	5.4e-65
patp:AAB42873 Human ORFX ORF2637 polypeptide sequence	664	5.4e-65
patp:AAG67449 Amino acid sequence of a human polypeptide	664	5.4e-65
patp:AAG67628 Amino acid sequence of a human protein	664	5.4e-65

In a search of public sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 347 of 420 bases (82%) identical to a gb:GENBANK-ID:HUMDSPHS|acc:L05147.1 mRNA from *Homo sapiens* (Human dual specificity phosphatase tyrosine/serine mRNA). The full amino acid sequence of the protein of the invention was found to have 135 of 185 amino acid residues (72%) identical to, and 150 of 185 amino acid residues (81%) similar to, the 185 amino acid residue ptnr:SWISSNEW-ACC:P51452 protein from *Homo sapiens* (Human) (DUAL SPECIFICITY PROTEIN PHOSPHATASE 3 (EC 3.1.3.48) (EC 3.1.3.16) (DUAL SPECIFICITY PROTEIN PHOSPHATASE VHR)). NOV74 also has homology to the other proteins shown in the BLASTP data in Table 74D.

Table 74D. NOV74 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17458321 ref XP_063782.1 (XM_063782)	similar to dual specificity phosphatase 3; vaccinia virus phosphatase Vh1-related; protein tyrosine phosphatase; serine/threonine specific protein phosphatase (H. sapiens) [Homo sapiens]	597	180/207 (86)	180/207 (86)	4e-92
gi 4758208 ref NP_004081.1 (NM_004090)	dual specificity phosphatase 3; vaccinia virus phosphatase Vh1-related; protein tyrosine phosphatase; serine/threonine specific protein phosphatase [Homo sapiens]	185	135/185 (72)	150/185 (80)	5e-71
gi 1633321 pdb 1VHR A	Chain A, Human Vh1-Related Dual-Specificity Phosphatase	184	137/184 (72)	149/184 (80)	2e-70
gi 18158941 pdb 1J4X A	Chain A, Human Vh1-Related Dual-Specificity Phosphatase C124s Mutant- Peptide Complex	184	133/184 (72)	148/184 (80)	3e-69
gi 12843112 dbj BAB25864.1 (AK008734)	DUAL SPECIFICITY PROTEIN PHOSPHATASE 3 (EC 3.1.3.48) (EC 3.1.3.16) (DUAL SPECIFICITY PROTEIN PHOSPHATASE VHR) ~putative [Mus musculus]	185	126/184 (68)	146/184 (78)	3e-65

This BLASTP data is displayed graphically in the ClustalW in Table 74E. A multiple sequence alignment is given, with the NOV74 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 74D.

Table 74E. ClustalW Alignment of NOV74	
NOV74	(SEQ ID NO:252)
gi 17458321	(SEQ ID NO:669)
gi 4758208	(SEQ ID NO:670)
gi 1633321	(SEQ ID NO:671)
gi 18158941	(SEQ ID NO:672)
gi 12843112	(SEQ ID NO:673)
<div> <div>1020304050</div> <div>..... </div> </div>	
NOV74	-----
gi 17458321	MGGQEETVEEIIQRMNSRSWVVRIVICCDFITITLRTEIWKPEQRGEALTQ
gi 4758208	-----
gi 1633321	-----
gi 18158941	-----
gi 12843112	-----

		60	70	80	90	100
NOV74					
gi 17458321	EFSSLSLCPMPGPGPADNKEYPRGQGRKAVPQPPLGTVWEQVCGGWSAR					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		110	120	130	140	150
NOV74					
gi 17458321	GTSCTKSQCTLDQERGERSGTHRONLSETSPNLPTAILGLLHPDLDVCKK					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		160	170	180	190	200
NOV74					
gi 17458321	KRCNRIYSGMEKTHPQALVRTQNGINEKNQKWDKRRCLGPGCHSAGLMP					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		210	220	230	240	250
NOV74					
gi 17458321	TAAYQLLAVLPGSPAPPDHSRLRGSEEVLAHTESTGDENMRHPQTPGLSKA					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		260	270	280	290	300
NOV74					
gi 17458321	LTAMQGAAREVGGHWELSPRLPRTSPGTSSSELSPHPLVPHPVHPPPINNQ					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		310	320	330	340	350
NOV74					
gi 17458321	LRERQDSRPAGKKAPVATWPFVSNLREKGPGLRRRGGSVPSIPDAAIMAIT					
gi 4758208	-----					
gi 1633321	-----					
gi 18158941	-----					
gi 12843112	-----					
		360	370	380	390	400
NOV74					
gi 17458321	-----MPGPTAAMTGFKKRSMEDLSDLLSDSGGCYSLPSQPCN					
gi 4758208	GALSASHCHLLNMPGPTAAMTGFKKRSMEDLSDLLSDSGGCYSLPSQPCN					
gi 1633321	-----MSGSFELSVQDLNDLLSDGSGGCYSLPSQPCN					
gi 18158941	-----SGSFELSVQDLNDLLSDGSGGCYSLPSQPCN					
gi 12843112	-----SGSFELSVQDLNDLLSDGSGGCYSLPSQPCN					
		410	420	430	440	450
NOV74					
gi 17458321	EVTLKIIYMGNTISVDQDIPKLOKLGSIHALNTTEGRSFMHIN-NANFSKDS					
gi 4758208	EVTLKIIYMGNTISVDQDIPKLOKLGSIHALNTTEGRSFMHIN-NANFSKDS					
gi 1633321	EVTPRIYVGNASVAQDIPKLOKLGITHVLNAAEGRSFMHVNTNANFYKDS					
		EVTPRIYVGNASVAQDIPKLOKLGITHVLNAAEGRSFMHVNTNANFYKDS				

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors. DUSP6 (alias PYST1), one of the dual-specificity tyrosine phosphatases, is localized on 12q21, one of the regions of frequent allelic loss in pancreatic cancer. This gene is composed of three exons, and two forms of alternatively spliced transcripts are ubiquitously expressed. Although no mutations were observed in 26 pancreatic cancer cell lines, reduced expressions of the full-length transcripts were observed in some cell lines, which may suggest some role for DUSP6 in pancreatic carcinogenesis. The mitogen-induced gene, DUSP2, encodes a nuclear protein, PAC1, that acts as a dual-specific protein phosphatase with stringent substrate specificity for MAP kinase. MAP kinase phosphorylation and consequent enzymatic activation is a central and often obligatory component in signal transduction initiated by growth factor stimulation or resulting from various types of oncogenic transformation. DUSP2 downregulates intracellular signal transduction through the dephosphorylation/inactivation of MAP kinases.

NOV74 is predicted to be expressed in at least the following tissues: retina. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV74 is provided in Example 2.

The NOV74 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. The NOV74 nucleic acid encoding the phosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 74A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 74A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 74A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 18% of the bases may be so changed.

The novel protein of the invention includes the dual specificity phosphatase-like protein whose sequence is provided in Table 74B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 74B while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 28% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV75

The disclosed NOV75 (alternatively referred to herein as CG56783-01) includes the 840 nucleotide sequence (SEQ ID NO:253) shown in Table 75A. A NOV75 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 31-33 and ends with a stop codon at nucleotides 769-771. The disclosed NOV75 maps to human chromosome 1.

Table 75A. NOV75 Nucleotide Sequence (SEQ ID NO:253)

AGAACCCAAGGCTCCCTGGATTGTCAGTCCATGAGCAACAAGCCCTGCTTACAGACCCCTGGGAGGAGCA
GATTCCACGAGGGGCTGGACCAAGTCTACCTTCCAATGTGGCTGGACTCTCTGCAGCCCCACCCAGAG
ACTGCCCATCAGGGAAGAGATGGTGCCATCAAGAGGCTATGGGGAAGAGGTGGATGAGGTCTGGCCCAAT
GTCTTCATAGCTGAGAAGAGTGTGGCTGTGAACAAGGGGAGGCTGAAGAGGCTGGGAATCACCACATTC
TGAATGCTGCGCATGGCACCAGCGTTTACACTGGCCCCGAATTCTACACTGGCCTGGAGATCCAGTACCT
GGGTGTAGAGGTGGATGACTTTCCTGAGGTGGACATTTCCAGCATTTCCGGAAGGCGTCTGAGTTCCTG
GATGAGGCGCTGCTGACTTACAGAGGACGTTTGACCAACGTGGGATTGAATGGGTCTGTGGTTCGCTGC
GGCGTAAGGAGTGTGTCCACCTCGCTCGCAGGTCTGGAGCGCACCGGCAGACCTCGAGGCGGAGCGGG
GAAAGTCTGGTCAGCAGCGAAATGGGCATCAGCCGGTCAGCAGTGTGGTGGTTCGCTACCTGATGATC
TTCACACATGGCCATCCTGGAGGCTTTGATGACCGTGCCTAAGAAGCGGCCATCTACCCCAATGAGG
GCTTCTGAAGCAGCTGCGGGAGCTCAATGAGAAGTTGATGGAGGAGAGAGAAGAGGACTATGCCCGGTA
GGGGGATCAGCTGAGGCTGAGGAGGGCGAGGGCACTGGGAGCATGCTCGGGCCAGAGTGCACGCCCTG

A NOV75 polypeptide (SEQ ID NO:254) encoded by SEQ ID NO:253 is 246 amino acids in length and is presented using the one-letter amino acid code in Table 75B. The Psort profile for NOV75 predicts that this sequence has no signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.7000. In alternative embodiments, a NOV75 polypeptide is located, to the endoplasmic reticulum (membrane) with a certainty of 0.1000, or to the nucleus with a certainty of 0.2000.

Table 75B. NOV75 Polypeptide Sequence (SEQ ID NO:254)

MSNKPCLQTPGRSRFHEGLDQVYLPNVAGLSAAPTQRLPIREEMVPSRGYGEDEVDEVPN
VFIAEKSVAVNKGRLKRLGITHILNAAHGTGVYTGPEFYTGLEIQYLGVEVDDFPPEVDIS
QHFRKASEFLDEALLTYRGRLTNVGLNGSVGLRRLRKECVPPRSQVLETRGRPRGGAGKVL
VSSEMGISRSVAVLVAYLMIFHNMAILEALMTVRKKRAIYPNEGFLKQLRELNEKLMEER
EEDYGR

A BLAST analysis of NOV75 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV75 had high homology to other proteins as shown in Table 75C.

Table 75C. BLASTX results from PatP database for NOV75

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum	
		Probability P (N)	
patp:AAE04836 Human SGP018 phosphatase polypeptide	389	1.6e-67	
patp:AAB40919 Human ORFX ORF683 polypeptide sequence	457	1.0e-55	
patp:AAE04837 Human SGP003 phosphatase polypeptide	218	2.2e-34	
patp:AAE04839 Human SGP060 phosphatase polypeptide	210	1.0e-32	
patp:AAE04839 Human SGP060 phosphatase polypeptide	210	1.0e-32	

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 167 of 257 bases (64%) identical to a gb:GENBANK-ID:AB027004|acc:AB027004.1 mRNA from *Homo sapiens* (mRNA for protein phosphatase).

The full amino acid sequence of the protein of the invention was found to have 39 of 89 amino acid residues (43%) identical to, and 58 of 89 amino acid residues (65%) similar to, the 198 amino acid residue ptnr:SPTREMBL-ACC:Q9QYJ7 protein from *Mus musculus* (Mouse) (PROTEIN PHOSPHATASE). NOV75 also has homology to the other proteins shown in the BLASTP data in Table 75D.

Table 75D. NOV75 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17454087 ref XP_061191.1 (XM_061191)	similar to protein phosphatase (H. sapiens) [<i>Homo sapiens</i>]	370	73/185 (39)	104/185 (55)	4e-30
gi 13128968 ref NP_076930.1 (NM_024025)	hypothetical protein MGC1136; hypothetical protein MGC2627 [<i>Homo sapiens</i>]	211	72/187 (38)	105/187 (55)	2e-29
gi 15072533 gb AAK77966.1 (AY040091)	branching-enzyme interacting dual-specificity protein phosphatase BEDP [<i>Homo sapiens</i>]	188	75/201 (37)	108/201 (53)	69e-29
gi 7305011 ref NP_038877.1 (NM_013849)	dual specificity phosphatase 13 [<i>Mus musculus</i>]	198	73/187 (39)	106/187 (56)	9e-29
gi 17390456 gb AAH18204.1 (AAH18204)(BC018204)	Similar to hypothetical protein MGC1136 [<i>Mus musculus</i>]	211	72/187 (38)	104/187 (55)	1e-28

This BLASTP data is displayed graphically in the ClustalW in Table 75E. A multiple sequence alignment is given, with the NOV75 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 75D.

Table 75E. ClustalW Alignment of NOV75	
NOV75	(SEQ ID NO:254)
gi 17454087	(SEQ ID NO:674)
gi 13128968	(SEQ ID NO:675)
gi 15072533	(SEQ ID NO:676)
gi 7305011	(SEQ ID NO:677)
gi 17390456	(SEQ ID NO:678)
	10 20 30 40 50
NOV75
	-----MSNKPCITQTPG-----R-----SR-----
gi 17454087	MLEVDPTVPVIMPNVKRDELEEVKEFAQVQRYGLGDVLFPLADGSLQTI
gi 13128968	-----MCPGNWLWASMTFMARFSR-----SSSR-----SPV-----
gi 15072533	-----MAETS-----
gi 7305011	-----MDSLQ-----X-----QEL-----
gi 17390456	-----MCPGNWLWASMTFMARFSR-----GSSR-----SPV-----

		60	70	80	90	100
NOV75					
gi 17454087	R	RG	PQ	AW	NV	NRLLGLPILGPKWPKPVGFSQI
gi 13128968	R	-----	-----	-----	-----	-----
gi 15072533	P	-----	-----	-----	-----	-----
gi 7305011	R	-----	-----	-----	-----	-----
gi 17390456	R	-----	-----	-----	-----	-----
		110	120	130	140	150
NOV75					
gi 17454087	S	LG	SK	MT	SG	EVKTS
gi 13128968	S	LG	SK	MT	SG	EVKTS
gi 15072533	S	LG	SK	MT	SG	EVKTS
gi 7305011	S	LG	SK	MT	SG	EVKTS
gi 17390456	S	LG	SK	MT	SG	EVKTS
		160	170	180	190	200
NOV75					
gi 17454087	P	SR	--	GY	GE	EV
gi 13128968	P	SR	--	GY	GE	EV
gi 15072533	P	SR	--	GY	GE	EV
gi 7305011	P	SR	--	GY	GE	EV
gi 17390456	P	SR	--	GY	GE	EV
		210	220	230	240	250
NOV75					
gi 17454087	Y	TG	PE	RY	TC	LE
gi 13128968	Y	TG	PE	RY	TC	LE
gi 15072533	Y	TG	PE	RY	TC	LE
gi 7305011	Y	TG	PE	RY	TC	LE
gi 17390456	Y	TG	PE	RY	TC	LE
		260	270	280	290	300
NOV75					
gi 17454087	N	VG	LN	GS	VG	RL
gi 13128968	N	VG	LN	GS	VG	RL
gi 15072533	N	VG	LN	GS	VG	RL
gi 7305011	N	VG	LN	GS	VG	RL
gi 17390456	N	VG	LN	GS	VG	RL
		310	320	330	340	350
NOV75					
gi 17454087	L	V	L	A	Y	L
gi 13128968	L	V	L	A	Y	L
gi 15072533	L	V	L	A	Y	L
gi 7305011	L	V	L	A	Y	L
gi 17390456	L	V	L	A	Y	L
		360	370	380	390	400
NOV75					
gi 17454087	S	Q	R	D	G	E
gi 13128968	S	Q	R	D	G	E
gi 15072533	S	Q	R	D	G	E
gi 7305011	S	Q	R	D	G	E
gi 17390456	S	Q	R	D	G	E
NOV75					
gi 17454087	R	N	Q	N	V	I
gi 13128968	R	N	Q	N	V	I
gi 15072533	R	N	Q	N	V	I

gi 7305011	-----
gi 17390456	-----

Table 75F lists the domain description from DOMAIN analysis results against NOV75. This indicates that the NOV75 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 75F. Domain Analysis of NOV75									
gnl Pfam pfam00782, DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region. SEQ ID NO:869									
CD-Length = 139 residues, 92.8% aligned Score = 98.2 bits (243), Expect = 5e-22									
NOV75:	56	EVWPNVFIAEKSVAVNKGRLKRLGITHILNAAHGTGVYTGPEFYTGLEIQYLGVEVDDFP	115						
		E+ P++++ A N L +LGITH++N F YL + VDD							
Sbjct:	4	EILPHLYLGSYPTASNLAFLSKLGITHVINVTEEVPSKNSGF-----LYLHIPVDDNH	57						
NOV75:	116	EVDISQHFRKASEFLDEALLTYRGRLTNVGLNGSVGRLRRKECVPPRSQVLERTGRPRGG	175						
		E DIS + +A EF+++A R							
Sbjct:	58	ETDISPYLDEAVEFIEDA-----RQK	78						
NOV75:	176	AGKVLVSSEMGISRSASVLVVAYLMIFHNMAILEALMTVRKKR-AIYPNEGFLKQ	228						
		GKVLV + GISRSA L++AYLM N+++ EA V+++R I PN GF +Q							
Sbjct:	79	GGKVLVHCQAGISRSATLI IAYLMKTRNLSLNEAYSFVKERRPIISPNGFKRQ	132						

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors. DUSP6 (alias PYST1), one of the dual-specificity tyrosine phosphatases, is localized on 12q21, one of the regions of frequent allelic loss in pancreatic cancer. This gene is composed of three exons, and two forms of alternatively spliced transcripts are ubiquitously expressed. Although no mutations were observed in 26 pancreatic cancer cell lines, reduced expressions of the full-length transcripts were observed in some cell lines, which may suggest some role for DUSP6 in pancreatic carcinogenesis. The mitogen-induced gene, DUSP2, encodes a nuclear protein, PAC1, that acts as a dual-specific protein phosphatase with stringent substrate specificity for MAP kinase. MAP kinase phosphorylation and consequent enzymatic activation is a central and often obligatory component in signal transduction initiated by growth factor stimulation or resulting

from various types of oncogenic transformation. DUSP2 downregulates intracellular signal transduction through the dephosphorylation/inactivation of MAP kinases.

NOV75 is predicted to be expressed in at least the following tissues: heart, breast and ovarian tissue, pancreas, brain, liver, kidney, spleen, testis, ovary, and peripheral blood leukocytes. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV75 is provided in Example 2.

The NOV75 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. The NOV75 nucleic acid encoding the phosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 75A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 75A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid.

The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 75A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such

CGTGCCACGCGTGACGCGTCGCTGTTAGAGGACAGCCGCTGGTACAGGCGCTCAGTGGGCTGCACCTG
TCCGCAGACAGGCTGGAAGACAGCAATAAGCTCAAGCGTTCCTTCTCTCTGGATATCAAATCAGTTTCAT
ATTCAGCCAGCATGGCAGCATCCTTACATGGCTTCTCCTCATCAGAAGATGCTTTGGAATACTACAAACC
TTCCACTACTCTGGATGGGACCAACAAGCTATGCCAGTTTCCCCTGTTTCAGGAAGTATCGGAGCAGACT
CCCGAAACAGTCCTGATAAGGAGGAAGCCAGCATCCCCAAGAAGCTGCAGACCGCCAGGCCTTCAGACA
GCCAGAGCAAGCGATTGCATTTCGGTCAGAACAGCAGCAGTGGCACCAGCCAGAGGTCCTTTTATCTCC
ACTGCATCGAAGTGGGAGCGTGGAGGACAATTACCACACCAGCTTCCTTTTCGGCCTTCCACAGCCAG
CAGCACCTCACGAAGTCTGCTGGCCTGGGCTTAAGGGCTGGCACTCGGATATCTTGGCCCCCAGACCT
CTACCCCTTCCCTGACCAGCAGCTGGTATTTTGCCACAGAGTCTCACACTTCTACTCTGCCTCAGCCAT
CTACGGAGGCACTGCCAGTTACTCTGCCTACAGCTGCAGCCAGCTGCCACTTGCGGAGACCAAGTCTAT
TCTGTGCGCAGGCGGCAGAACCAAGTGACAGAGCTGACTCGCGGCGGAGCTGGCATGAAGAGAGCCCT
TTGAAAAGCAGTTTAAACGCAGAAGCTGCCAAATGGAATTTGGAGAGAGCATCATGTACAGAAACAGGTC
ACGGGAAGAGCTGGGGAAGTGGGCAGTCACTAGCTTTTCGGGCAGCATGGAAATCATTGAGGTCTCC
TGAGAAGAAAGACACTGTGACTTCTATAGACAATTTTTTTTCTGTTCCAAAAAAATTCCTGTAA
TCTGAAATATATATATGTACATACATAT

5 A NOV76a polypeptide (SEQ ID NO:256) encoded by SEQ ID NO:255 is 680 amino acids in length and is presented using the one-letter amino acid code in Table 76B. The Psort profile for NOV76a predicts that this sequence has no signal peptide and is likely to be localized at the nucleus with a certainty of 0.8800. In alternative embodiments, a NOV76a polypeptide is located to peroxisomal microbodies with a certainty of 0.3000, to the mitochondrial matrix space with a certainty of 0.1000, or to lysosomes with a certainty of 0.1000.

Table 76B. NOV76a Polypeptide Sequence (SEQ ID NO:256)
MAHEMIGTQIVTERLVALLESgtekvLLIDSRPFVEYNTSHILEAININCSKLMKRRLQQ DKVLITELIQHSAKHKVNAQVDIDCSQKVVDYDQSSQDVASLSSDCFLTVLGKLEKSFN SVHLLAGLFLGGFAEFRCFPGLCEGKSTLVPTCISQCLPVANIGPTRILPNLYLGCQR DVLNKLMMQNGIGYVLNASNTCPKPDFIPESHFLRVPVNDSECEKILPWLDKSVDFIGK LTYTEKAKASNGCVLVHCLAGISRSATIAIAYIMKRMDMSLDEAYRRFVKEKRPTISPNF NFLGQLLDYEKKIKNQGTGASGPKSKLKLHLEKPNPVPVAVSEGGQKSETPLSPPCADSA TSEAAGQRPVHPASVPSVPSVQPSLLEDSPLVQALSGHLHLSADRLEDNKLKRSFSLDIK SVSYSASMAASLHGFSSSEDALEYKPSSTLDTGNKLCQFSPVQELSEQTPETSPDKKEA SIPKKLQTARPSDSQSKRLHSVRTSSSGTAQRSLLSPLHRSGSVEDNYHTSFLFGLSTSQ QHLLKSAAGLGLKGWHSIDILAPQSTPSLTSSWYFATESSHFYASAIYGGASAYSAYSCS QLPTCGDQVYSVRRRQKPSDRADSRRSWHEESPFEKQFKRRSCQMEFGESIMSENRSREE LGKVGSSQSFSGSMEEIEVS

10

NOV76b

The disclosed NOV76b (alternatively referred to herein as CG56789-02) includes the 2071 nucleotide sequence (SEQ ID NO:257) shown in Table 76C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 61-63 and ends with a stop codon at 15 nucleotides 2047-2049. The disclosed NOV76b maps to human chromosome 12.

Table 76C. NOV76b Nucleotide Sequence (SEQ ID NO:257)
ACCATTACATCATCGTGGCAAATTAAGAAGGAGGTGGGAAAAGAGGACTTATTGTGTGTC ATGGCCCATGAGATGATTGGAACCTCAAATTGTTACTGAGAGGTTGGTGGCTCTGCTGGAA AGTGAACGGAAAAAGTGTGCTAATTGATAGCCGCCATTGTGGAATACAATACATCC

CACATTTTGAAGCCATTAATATCAACTGCTCCAAGCTTATGAAGCGAAGGTTGCAACAG
GACAAAGTGTTAATTACAGAGCTCATCCAGCATTGAGCGAAACATAAGGTTGACATTGAT
TGCAGTCAGAAGGTTGTAGTTTACGATCAAAGCTCCCAAGATGTTGCCCTCTCTCTTCA
GACTGTTTTCTCACTGTACTTCTGGGTAAACTGGAGAAGAGCTTCAACTCTGTTCACCTG
CTTGCAGGTGGGTTTGTGAGTTCTCTCGTTGTTTCCCTGGCCTCTGTGAAGGAAATCC
ACTCTAGTCCCTACCTGCATTTCTCAGCCTTGCTTACCTGTGCGCAACATTGGGCCAAC
CGAATTCTTCCCAATCTTTATCTTGGCTGCCAGCGAGATGTCCTCAACAAGGAGCTGATG
CAGCAGAATGGGATTGGTTATGTGTTAAATGCCAGCAATACCTGTCCAAGCCTGACTTT
ATCCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAAATTTG
CCGTGGTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAGCCTCCAATGGATGTGTT
CTAGTGCATGTTTAGCTGGGATCTCCGCTCCGCCACCATCGCTATCGCCTACATCATG
AAGAGGATGGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACCTACT
ATATCTCCAAACTTCAATTTCTGGGCCAACTCTGGACTATGAGAAGAAGATTAAAGAAC
CAGACTGGAGCATCAGGGCCAAAGAGCAAACCTCAAGCTGCTGCACCTGGAGAAGCCAAAT
GAACCTGTCCCTGTCTCTCAGAGGTTGGACAGAAAAGCGAGACGCCCCCTCAGTCCACCC
TGTGCCGACTCTGCTACCTCAGAGGCAGCAGGACAAAGGCCGTGCATCCCGCCAGCGTA
CCCAGCGTGACGCGTCTGCTGTAGAGGACAGCCGCTGGTACAGGCGCTCAGTGGGCTG
CACCTGTCCGACAGAGGCTGGAAGACAGCAATAAGCTCAAGCGTTCTCTCTCTGAT
ATCAAATCAGTTTCATATTAGCCAGCATGGCAGCATCCTTACATGGCTTCTCTCATCA
GAAGATGCTTTGGAATACTACAAACCTTCCACTACTCTGGATGGGACCAACAAGCTATGC
CAGTCTCTCCCTGTTTCCAGGAATATCGGAGCAGACTCCCGAAACCAGTCTGATAAGGAG
GAAGCCAGCATCCCCAAGAAGCTGACAGCCGCCAGGCGCTTCCAGACGCCAGAGCAAGCGA
TTGCATTGCGTCCAGAACAGCAGCAGTGGCACCGCCAGAGTCCCTTTTATCTCCACTG
CATCGAAGTGGGAGCGTGGAGGACAATTACCACACAGCTTCTTTTCGGCCTTTCCACC
AGCCAGCAGCACCTCACGAAGTCTGCTGGCCTGGGCTTAAGGGCTGGCACTCGGATATC
TTGGCCCCCAGACCTCTACCCCTTCCCTGACCAGCAGCTGGTATTTTGGCCAGAGTCC
TCACACTTCTACTCTGCCTCAGCCATCTACGGAGGAGTGCAGTTACTCTGCCTACAGC
TGCAGCCAGCTGCCCACTTGCAGGAGACCAAGTCTATTCTGTGCGCAGGCGGCAGAACCA
AGTGACAGAGCTGACTCGCGGGGAGCTGGCATGAAGAGAGCCCTTTGAAAAGCAGTTT
AAACGCAGAAGCTGCCAAATGGAATTTGGAGAGAGCATCATGTGAGAGAACAGGTCACGG
GAAGAGCTGGGAAAGTGGGCAGTCACTGCTTTTCGGGCAGCATGGAATCATTGAG
GTCTCTGAGAAGAAAGACACTTGTGACTTC

A NOV76b polypeptide (SEQ ID NO:258) encoded by SEQ ID NO:257 is 662 amino acids in length and is presented using the one-letter amino acid code in Table 76D. The Psort profile for NOV76b predicts that this sequence has no signal peptide and is likely to be localized to the nucleus with a certainty of 0.8800. In alternative embodiments, a NOV76b polypeptide is located to peroxisomal microbodies with a certainty of 0.3000.

Table 76D. NOV76b Polypeptide Sequence (SEQ ID NO:258)

MAHEMIGTQIVTERLVALLESgtekvLLIDSRPFVEYNTSHILEAININCSKLMKRRLQQ
DKVLITELIQHSAKHVDIDCSQKVVDQSSQDVASLSSDCFITVLLGKLEKSFNSVHL
LAGGFAEFRCFPGLCEGKSTLVPTCISQPCLPVANIGPTRILPNLYLGCQRDVLNKELM
QQNGIGYVLNASNTCPKPDFIPESHFLRVPVNDSECEKILPWLDKSVDIEKAKASNGCV
LVHCLAGISRSATIAIAYIMKRMDMSLDEAYRFVKEKRPTISPNFNFLGQLLDYEKKIKN
QTGASGPKSKLLHLLEKPNPEVPAVSEGGQKSETPLSPCADSATSEAAGQRPVHPASV
PSVQPSLLEDSPVLQALSGLHLSADRLSDSNLKRSLDIKSVSYASMAASLHGFSSS
EDALEYYKPSTTLDGTNKLQCFSPVQELSEQTPETSPDKEEASIPKKLQATARPSDSQSKR
LHSVRTSSSGTAQRSLLSPLHRSGSVEDNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSDI
LAPQTSTPSLTSSWYFATESSHFYASAIYGGASAYSAYSCSCLPTCGDQVYSVRRRKP
SDRADSRRSWHESPFQKRRSCQMEFGESIMSENRSREELGKVGSQSSFSGSMETIE
VS

A BLAST analysis of NOV76 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV76 had high homology to other proteins as shown in Table 76E.

Table 76E. BLASTX results from PatP database for NOV76

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAE04834 Human SGP002 phosphatase polypeptide	3360	0.0
patp:AAU09016 Human dual specificity phosphatase 21117	3360	0.0
patp:AAB20325 Human protein phosphatase and kinase protein	2963	1.3e-308
patp:AAM25744 Human protein sequence	2860	1.1e-297
patp:AAW29150 Dual-specific murine thr-tyr phosphatase	1088	1.5e-125

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1149 of 1150 bases (99%) identical to a gb:GENBANK-
5 ID:AB052156|acc:AB052156.1 mRNA from *Homo sapiens* (MKP-7 mRNA for MAPK phosphatase-7). The full amino acid sequence of the protein of the invention was found to have 662 of 665 amino acid residues (99%) identical to, and 662 of 665 amino acid residues (99%) similar to, the 665 amino acid residue ptnr:SPTREMBL-ACC:Q9BY84 protein from *Homo sapiens* (Human) (MAPK PHOSPHATASE-7). NOV76 also has homology to the
10 other proteins shown in the BLASTP data in Table 76F.

Table 76F. NOV76 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 12697945 d bj BAB21791.1 (AB051487)	KIAA1700 protein [<i>Homo sapiens</i>]	690	665/680 (97)	665/680 (97)	0.0
gi 14756395 f ef XP_039106. 1 (XM_039106)	MAPK phosphatase-7 [<i>Homo sapiens</i>]	665	665/680 (97)	665/680 (97)	0.0
gi 16550836 d bj BAB71060.1 (AK055973)	unnamed protein product [<i>Homo sapiens</i>]	665	664/680 (97)	664/680 (97)	0.0
gi 13990989 d bj BAB47240.1 (AB052157)	MAP kinase phosphatase-7 [<i>Mus musculus</i>]	660	601/680 (88)	628/680 (91)	0.0
gi 13625393 g b AAK35052.1 AF345951_1 (AF345951)	map kinase phosphatase-M A1 isoform [<i>Mus musculus</i>]	677	533/644 (82)	568/644 (87)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 76G. A multiple sequence alignment is given, with the NOV76 protein being shown on line 1 in a ClustalW
15 analysis comparing the protein of the invention with the related protein sequences shown in Table 76F.

gi 13625393	KILPWLDKSVDFI-----EKAKASNGCVLHCLAGISRSATIAIAYIMK
	310 320 330 340 350
NOV76a	RMDMSLDEAYRRFVKEKRPTISPNFNFLGQLLDYEKKIKNOTGASGPKSK
NOV76b	RMDMSLDEAYR-FVKEKRPTISPNFNFLGQLLDYEKKIKNOTGASGPKSK
gi 12697945	RMDMSLDEAYR-FVKEKRPTISPNFNFLGQLLDYEKKIKNOTGASGPKSK
gi 14756395	RMDMSLDEAYR-FVKEKRPTISPNFNFLGQLLDYEKKIKNOTGASGPKSK
gi 16550836	RMDMSLDEAYR-FVKEKRPTISPNFNFLGQLLDYEKKIKNOTGASGPKSK
gi 13990989	RMDMSLDEAYR-FVKEKRPTISPNFNFMGQLMDYEKTIINNOTGMSGPKSK
gi 13625393	RMDMSLDEAYR-FVKEKRPTISPNFNFMGQLMDYEKTIINNOTGMSGPKSK
	360 370 380 390 400
NOV76a	LKLLHLEKPNPVPVAVSEGGQKSETPLSPPCADSATSEAGQRPVHPASV
NOV76b	LKLLHLEKPNPVPVAVSEGGQKSETPLSPPCADSATSEAGQRPVHPASV
gi 12697945	LKLLHLEKPNPVPVAVSEGGQKSETPLSPPCADSATSEAGQRPVHPASV
gi 14756395	LKLLHLEKPNPVPVAVSEGGQKSETPLSPPCADSATSEAGQRPVHPASV
gi 16550836	LKLLHLEKPNPVPVAVSEGGQKSETPLSPPCADSATSEAGQRPVHPASV
gi 13990989	LKLLHLDKPSVPVPAASEGGWKSALSLSPPCAN-STSEASGQRLVHPASV
gi 13625393	LKLLHLDKPSVPVPAASEGGWKSALSLSPPCAN-STSEASGQRLVHPASV
	410 420 430 440 450
NOV76a	PSVPSVQPSLLEDSPVLQALSGLHLSADRLEDSENKLRFSFLDIKSVSYS
NOV76b	P---SVQPSLLEDSPVLQALSGLHLSADRLEDSENKLRFSFLDIKSVSYS
gi 12697945	PSVPSVQPSLLEDSPVLQALSGLHLSADRLEDSENKLRFSFLDIKSVSYS
gi 14756395	PSVPSVQPSLLEDSPVLQALSGLHLSADRLEDSENKLRFSFLDIKSVSYS
gi 16550836	PSVPSVQPSLLEDSPVLQALSGLHLSADRLEDSENKLRFSFLDIKSVSYS
gi 13990989	P---SLQPSLLEDSPVLQALSGLQLSSEKLEDSTKLKRSFSLDIKSVSYS
gi 13625393	P---SLQPSLLEDSPVLQALSGLQLSSEKLEDSTKLKRSFSLDIKSVSYS
	460 470 480 490 500
NOV76a	ASMAASLHGFSSSEDALEYKPTSTLDGTNKLQCFSPVQELSEQTPETSP
NOV76b	ASMAASLHGFSSSEDALEYKPTSTLDGTNKLQCFSPVQELSEQTPETSP
gi 12697945	ASMAASLHGFSSSEDALEYKPTSTLDGTNKLQCFSPVQELSEQTPETSP
gi 14756395	ASMAASLHGFSSSEDALEYKPTSTLDGTNKLQCFSPVQELSEQTPETSP
gi 16550836	ASMAASLHGFSSSEDALEYKPTSTLDGTNKLQCFSPVQELSEQTPETSP
gi 13990989	ASMAASLHGFSS-EEALDYCKPSATLDGTNKLQCFSPVQELSEQTPETSP
gi 13625393	ASMAASLHGFSS-EEALDYCKPSATLDGTNKLQCFSPVQELSEQTPETSP
	510 520 530 540 550
NOV76a	DKEEASIPKKLQTARPSDSQSKRLHSVRTSSSGTAQSRLLSPLHRSGSVE
NOV76b	DKEEASIPKKLQTARPSDSQSKRLHSVRTSSSGTAQSRLLSPLHRSGSVE
gi 12697945	DKEEASIPKKLQTARPSDSQSKRLHSVRTSSSGTAQSRLLSPLHRSGSVE
gi 14756395	DKEEASIPKKLQTARPSDSQSKRLHSVRTSSSGTAQSRLLSPLHRSGSVE
gi 16550836	DKEEASIPKKLQTARPSDSQSKRLHSVRTSSSGTAQSRLLSPLHRSGSVE
gi 13990989	DKEEAHPIKQPQPPRPSESQVTRLHSVRTCSSGSTQRPFFSPLHRSGSVE
gi 13625393	DKEEAHPIKQPQPPRPSESQVTRLHSVRTGSSGSTQRPFFSPLHRSGSVE
	560 570 580 590 600
NOV76a	DNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSIDILAPQTSTPSLTSSWYFA
NOV76b	DNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSIDILAPQTSTPSLTSSWYFA
gi 12697945	DNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSIDILAPQTSTPSLTSSWYFA
gi 14756395	DNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSIDILAPQTSTPSLTSSWYFA
gi 16550836	DNYHTSFLFGLSTSQQHLTKSAGLGLKGWHSIDILAPQTSTPSLTSSWYFA
gi 13990989	DNYHTNFLFGLSTSQQRLTKSAGLGLKGWHSIDILAPQSSAPSLTSSWYFA
gi 13625393	DNYHTNFLFGPFHQPATPHQVCRAWLKGWHSIDILAPQSSAPSLTSSWYFA
	610 620 630 640 650
NOV76a	TESSHFYASAIYGGASAYSAYSCSQLPTCGDQVYSVRRRQKPSDRADSR
NOV76b	TESSHFYASAIYGGASAYSAYSCSQLPTCGDQVYSVRRRQKPSDRADSR
gi 12697945	TESSHFYASAIYGGASAYSAYSCSQLPTCGDQVYSVRRRQKPSDRADSR
gi 14756395	TESSHFYASAIYGGASAYSAYSCSQLPTCGDQVYSVRRRQKPSDRADSR
gi 16550836	TESSHFYASAIYGGASAYSAYSCSQLPTCGDQVYSVRRRQKPSDRADSR

gi 13990989	TEPSHLYSASAIYGGNSYSAYSCGQLPTCSDOTYSVRRRQKPTDRADSR
gi 13625393	TEPSHLYSASAIYGGNSYSAYSCGQLPTCSDOTYSVRRRQKPTDRADSR
	660 670 680 690 700
NOV76a	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
NOV76b	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
gi 12697945	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
gi 14756395	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
gi 16550836	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
gi 13990989	RSWHEESPFEKQFKRRSCQMEFGESIMSENRSREELG---KVGSQSSFS
gi 13625393	RTGMKRAPLKSSSLNAEAAKWNLERALCRRTGPGRSWARWAASPAAPAWR
	710 720
NOV76a	GSMETIEVS-----
NOV76b	GSMETIEVS-----
gi 12697945	GSMETIEVS-----
gi 14756395	GSMETIEVS-----
gi 16550836	GSMETIEVS-----
gi 13990989	GSMETIEVS-----
gi 13625393	SSRSLEKTS SLLLTVLFPVHKK

Table 76H lists the domain description from DOMAIN analysis results against NOV76. This indicates that the NOV76 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 76H. Domain Analysis of NOV76					
gnl Pfam pfam00782, DSPc, Dual specificity phosphatase, catalytic domain.					
Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region. SEQ ID NO:870					
CD-Length = 139 residues, 100.0% aligned					
Score = 172 bits (436), Expect = 6e-44					
NOV76:	166	GPTRILPNLYLGCQRDVLNKELMQQNGIGYVLNASNTCPKPDFIPESHFLRVPVNDSE	225		
		GP+ ILP+LYLG N + + GI +V+N + P +L +PV+D+			
Sbjct:	1	GPSEILPHLYLGSYPTASNLAFLSKLGITHVINVTVEEVPN-SKNSGFLYLHIPVDDNHET	59		

NOV76:	226	KILPWLDKSVDFIGKLTYTEKAKASNGCVLVHCLAGISRSATIAIAYIMKRMDMSLDEAY	285		
		I P+LD++V+FI E A+ G VLVHC AGISRSAT+ IAY+MK ++SL+EAY			
Sbjct:	60	DISPYLDEAVEFI-----EDARQKGGKVLVHCQAGISRSATLIIAYLMKTRNLSLNEAY	113		
.					
NOV76:	286	RRFVKEKRPTISPNFNFLGQLLDYEKK	312		
		FVKE+RP ISPNF F QL++YE+K			
Sbjct:	114	-SFVKERRPIISPNFGFKROLIEYERK	139		

Mitogen-activated protein kinases (MAPKs) are inactivated via dephosphorylation of either the threonine or tyrosine residue or both in the P-loop catalyzed by protein phosphatases which include serine/threonine phosphatases, tyrosine phosphatases, and dual specificity phosphatases. Nine members of the dual specificity phosphatases specific for MAPKs, termed MKPs, have been reported. Each member has its own substrate specificity, tissue distribution,

and subcellular localization. MKP-7 is most similar to hVH5, a member of previously known MKPs, in the primary structure. MKP-7 is predominantly localized in the cytoplasm when expressed in cultured cells, whereas hVH5 is both in the nucleus and the cytoplasm. MKP-7 binds to and inactivates p38 MAPK and JNK/SAPK, but not ERK. Furthermore, MKPs have the substrate specificity toward the isoforms of the p38 family (alpha, beta, gamma, and delta). MKP-7 binds to and inactivates p38 alpha and -beta, but not gamma or delta. MKP-5 and CL100/MKP-1 also bind to p38 alpha and -beta, but not gamma or delta.

NOV76 is predicted to be expressed in at least the following tissues: blood, brain, CNS, colon, heart, kidney, lung, and stomach. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV76 is provided in Example 2.

The NOV76 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, hypercalcaemia, ulcers, Hirschsprung's disease, Crohn's Disease, anemia, ataxia-telangiectasia, autoimmune disease, immunodeficiencies, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neurodegeneration, neuroprotection, systemic lupus erythematosus, asthma, emphysema, allergy, ARDS, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, renal tubular acidosis, IgA nephropathy, as well as other diseases, disorders and conditions. NOV76 nucleic acids encoding the MAP kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 76A or 76C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 76A or 76C while still encoding a protein that maintains its phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary

to the sequence of Table 76A or 76C, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the phosphatase-like protein whose sequence is provided in Table 76B or 76D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 76B or 76D while still encoding a protein that maintains its phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV77

The disclosed NOV77 (alternatively referred to herein as CG56804-01) includes the 881 nucleotide sequence (SEQ ID NO:259) shown in Table 77A. A NOV77 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 61-63 and ends with a stop codon at nucleotides 769-771. The disclosed NOV77 maps to human chromosome 14.

Table 77A. NOV77 Nucleotide Sequence (SEQ ID NO:259)

```
GAGGGTCGGCCGGCTGTGTAACACTCTCCACCCCCACCCAGCCCCGGGGCCAGCACCATTGGAGGACG
TGAAGCTGGAGTTCCCTTCCCTTCCACAGTGCAAGGAAGACGCCGAGGAGTGGACCTACCTGAGTGGAC
CTACCCCTATGAGACGAGAGATGCAGGAAATTTTACCTGGATTGTTCTTAGGCCCATATTCATCTGCTATG
AAAAGCAAGGTACTACCTGTACTACAGAAACATGGAATAACCCATATAATATGCATACGACAAAATATTG
AAGCAAACCTTTATTAAACCAAACCTTTCAGCAGTTATTTAGGTATTTAGTCCTGGATATTGCAGATAATCC
AGTTGAAAATATAATACGTTTTTCCCTATGTTTTGCCTCCAGACTAAGGAATTTATGATGGGAGCTTA
CAAATGGGAGGTAAAGTTCTTGTGCATGGAAATGCAGGGATCTCCAGAAGTCAGCCTTTGTTATTGCAT
ACATTATGGAACATTTGGAATGAAGTACAGGTTTCAGAGATGCTTTTGCTTATGTTCAAGAAAGAAGATT
```

TTGTATTAACTCCTAATGCTGGATTGTCCATCAACTTCAGGAATATGAAGCCATCTACCTAGCAAAATTA
ACAATACAGATGATGTCACCACTCCAGATAGAAAGGTCATTATCTGTTTCATTCTGGTACCACAGGTGGCA
GTTTGAAGAGAACACATGAAGAAGAGGATGATTTTGAACCATGCAAGTGGCGACTGCACAGAATGGCTG
ACTTGAAGAGCAACATCATAGAGTGTGAATTTCTATTGTTGGGAAGGAGAAAATACAAGAGAAAATTATAAT
GTAAATGGTAAAAACATAAGTAGTTTTTTTTTCAATTACA

A NOV77 polypeptide (SEQ ID NO:260) encoded by SEQ ID NO:259 is 236 amino acids in length and is presented using the one-letter amino acid code in Table 77B. The Psort profile for NOV77 predicts that this sequence has no signal peptide and is likely to be localized at the cytoplasm with a certainty of 0.6036. In alternative embodiments, a NOV77 polypeptide is located to lysosomes with a certainty of 0.2040.

Table 77B. NOV77 Polypeptide Sequence (SEQ ID NO:260)

MEDVKLEFPSPQCKEDAEEWTYPEWTYPEPMRREMQEILPGLFLGPYSSAMKSKVLPVLQK
HGITHIICIRQNIENFIKPNFQQLFRYLVLADIADNPVENIIRFFPMFCLQTKFIDGSL
QMGKVLVHGNAGISRSAFVIAYIMETFGMKYRFRDAFAYVQERRFCINPNAGFVHQLQ
EYEAIYLAKLTIQMSPLQIERSLSVHSGTTGGSLKRTHEEEDDFGTMQVATAQNG

A BLAST analysis of NOV77 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV77 had high homology to other proteins as shown in Table 77C.

Table 77C. BLASTX results from PatP database for NOV77

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAM39734 Human polypeptide	1099	4.3e-111
patp:AAM41520 Human polypeptide	1099	4.3e-111
patp:AAE08552 Human phosphatase protein - <i>Homo sapiens</i>	1099	4.3e-111
patp:AAU09017 Human dual specificity phosphatase 38692	1099	4.3e-111
patp:AAV68795 Amino acid sequence of a human protein	210	6.9e-17

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 228 of 249 bases (91%) identical to a gb:GENBANK-ID:MMU34973|acc:U34973.1 mRNA from *Mus musculus* (protein tyrosine phosphatase-like mRNA, unspliced c-terminal product and spliced c-terminal end STYX). The full amino acid sequence of the protein of the invention was found to have 214 of 236 amino acid residues (90%) identical to, and 221 of 236 amino acid residues (93%) similar to, the 223 amino acid residue ptrn:SPTREMBL-ACC:Q60970 protein from *Mus musculus* (Mouse) (PROTEIN TYROSINE PHOSPHATASE-LIKE). NOV77 also has homology to the other proteins shown in the BLASTP data in Table 77D.

Table 77D. NOV77 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17476793 ref XP_058659.1 (XM_058659)	similar to putative (H. sapiens) [<i>Homo sapiens</i>]	223	223/236 (94)	223/236 (94)	e-118
gi 12833088 dbj BAB22384.1 (AK002822)	phosphoserine/threonine/tyrosine interaction protein-putative [<i>Mus musculus</i>]	223	215/236 (91)	221/236 (93)	e-116
gi 2137698 pir I49365	protein tyrosine phosphatase - mouse	233	214/236 (90)	221/236 (92)	e-116
gi 9789981 ref NP_062611.1 (NM_019637)	phosphoserine/threonine/tyrosine interaction protein; STNS (alternatively spliced intron of Styx); protein tyrosine phosphatase-like unspliced c-terminal product and spliced c-terminal end STYX [<i>Mus musculus</i>]	205	163/180 (90)	167/180 (92)	3e-87
gi 1842088 gb AAB47561.1 (U87169)	tyrosine phosphatase-like protein homolog hSTYXb [<i>Homo sapiens</i>]	66	66/68 (97)	66/68 (97)	8e-30

This BLASTP data is displayed graphically in the ClustalW in Table 77E. A multiple sequence alignment is given, with the NOV77 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 77D.

Table 77E. ClustalW Alignment of NOV77	
NOV77	(SEQ ID NO:260)
gi 17476793	(SEQ ID NO:684)
gi 12833088	(SEQ ID NO:685)
gi 2137698	(SEQ ID NO:686)
gi 9789981	(SEQ ID NO:687)
gi 1842088	(SEQ ID NO:688)
	10 20 30 40 50
NOV77	MEDVKLEFPSPQCKEDAEEWTPPEWTPMRREMQEILPGLFLGPYSSAM
gi 17476793	MEDVKLEFPSPQCKEDAEEWTPPEWTPMRREMQEILPGLFLGPYSSAM
gi 12833088	MEDVKLEFPSPQCKEDAEEWTPPEWTPMRREMQEILPGLFLGPYSSAM
gi 2137698	MEDVKLEFPSPQCKEDAEEWTPPEWTPMRREMQEILPGLFLGPYSSAM
gi 9789981	MEDVKLEFPSPQCKEDAEEWTPPEWTPMRREMQEILPGLFLGPYSSAM
gi 1842088	-----
	60 70 80 90 100
NOV77	KSKVLPVQLQKHGITHIICIRONIEANFIKPNFQQLFRYLVLADIADNPVEN
gi 17476793	KSK-LPVQLQKHGITHIICIRONIEANFIKPNFQQLFRYLVLADIADNPVEN
gi 12833088	KSK-LPVQLQKHGITHIICIRONIEANFIKPNFQQLFRYLVLADIADNPVEN
gi 2137698	KSK-LPVQLQKHGITHIICIRONIEANFIKPNFQQLFRYLVLADIADNPVEN
gi 9789981	KSK-LPVQLQKHGITHIICIRONIEANFIKPNFQQLFRYLVLADIADNPVEN

gi 1842088	-----
	110 120 130 140 150
NOV77	IIRFFPMFCLQTKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFG
gi 17476793	IIRFFPM---TKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFG
gi 12833088	IIRFFPM---TKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFG
gi 2137698	IIRFFPM---TKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFG
gi 9789981	IIRFFPM---TKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFG
gi 1842088	-----VLVHGNAGISRSAAFVIAYIMETFG
	160 170 180 190 200
NOV77	MKYRFRDAFAYVQERRFCINPNAGFVHQLQEYEAIIYLAKLTIQMMSPLQI
gi 17476793	MKY--RDAFAYVQERRFCINPNAGFVHQLQEYEAIIYLAKLTIQMMSPLQI
gi 12833088	MKY--RDAFAYVQERRFCINPNAGFVHQLQEYEAIIYLAKLTIQMMSPLQI
gi 2137698	MKY--RDAFAYVQERRFCINPNAGFVHQLQEYEAIIYLAKLTIQMMSPLQI
gi 9789981	MKY--RDAFAYVQERRFCINPNAGFVHQLQLWLSWNSARS-----APLPI
gi 1842088	MKY--RDAFAYVQERRFCINPNAGFVHQLQEYEAIIYLAKLTIQ-----
	210 220 230
NOV77	ERSLSVHSGTTGCSLKRTHEEEDDFGTMQVATAQNG
gi 17476793	ERSLSVHSGTTG-SLKRTHEEEDDFGTMQVATAQNG
gi 12833088	ERSLAVHSGTTG-SVKRTHEEEDDFGTMQVATAQNG
gi 2137698	ERSLAVHSGTTG-SVKRTHEEEDDFGTMQVATAQNG
gi 9789981	KQRQVYHCAFKT--SK-NKQTNNSE-----
gi 1842088	-----

Table 77F lists the domain description from DOMAIN analysis results against NOV77. This indicates that the NOV77 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 77F. Domain Analysis of NOV77				
gnl Smart smart00195, DSPc, Dual specificity phosphatase, catalytic domain				
SEQ ID NO:871				
CD-Length = 139 residues, 98.6% aligned				
Score = 142 bits (358), Expect = 2e-35				
NOV77:	33	EMQEILPGLFLGPYSSAMKSKVLPVLQKHGITHIICIRQNIEANFIKPNFQQFLFRYLVD	92	
		EILP L+LG YS A L +L+K GITH+I + + + F YL +		
Sbjct:	1	GPSEILPHLYLGSYSASN---LALLKKLGITHVINV---TEEVPSNKSGLFLYLGLIP	52	
NOV77:	93	IADNPVENIIRFFPMFCLQTKEFIDGSLQMGKVLVHGNAGISRSAAFVIAYIMETFGMK	152	
		+ DN I + P EFI+ + + GGKVLVH AG+SRSA +IAY+M+ M		
Sbjct:	53	VDDNTETKISPYLPE---AVEFIEDAEKKGGKVLVHCQAGVSRSATLIIAYLMKYRNMS	108	
NOV77:	153	YRFRDAFAYVQERRFCINPNAGFVHQLQEYE	183	
		DA+ +V+ERR I+PN GF+ QL EYE		
Sbjct:	109	L---NDAYDFVKERRPIISPNGFLRLQIEYE	137	

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular

processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors.

NOV77 is predicted to be expressed in at least the following tissues: lung, lymphoid tissue, spleen, tonsils, whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV77 is provided in Example 2.

The NOV77 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. The NOV77 nucleic acid encoding the MAP kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 76A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 76A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 76A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In the mutant or variant nucleic acids, and their complements, up to about 9% of the bases may be so changed.

The novel protein of the invention includes the dual specificity phosphatase-like protein whose sequence is provided in Table 76B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 76B while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 10% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV78

The disclosed NOV78 (alternatively referred to herein as CG56810-01) includes the 777 nucleotide sequence (SEQ ID NO:261) shown in Table 78A. A NOV78 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 61-63 and ends with a stop codon at nucleotides 768-770. The disclosed NOV78 maps to human chromosome 2.

Table 78A. NOV78 Nucleotide Sequence (SEQ ID NO:261)

```
TCACCTGAGCCTAGGAGTTCAAGATTGCAGTGGCCTATGATTGCATCATTGCACTCCAGCCTGGGTGACA
TCACCTGAGCCTAGGAGTTCAAGATTGCAGTGGCCTATGATTGCATCATTGCACTCCAGCCTGGGTGACA
GAAAGAGACCCCTGTCTCTGCATGATGATAATAATAATTTAAAGAGAGAGAGAGAGAGAGAAAACATATAG
TAGAAGGAGCACTTCTGGTGTGGGATTGAAGCAGTATTATCATTCAAGTGAGTCGCATTAAATTTTTA
CCTTTCTTGTTACCGATATCAACACCCCACTCTCTTTTAGGTACCCCTAATTCATGGGATTTAAACAAT
CATCTTTTTCTTTCTTTAAGGTGACTCATATTCTTAATGTTGCATATGGAGTTGAAAATGCTTTCTT
CAGTGACTTTACATATAAGAGCATTCTATATTGGATCTGCCTGAAACCAACATCCTGTCTTATTTCCA
GAATGTTTTGAATTTATTGAAGAAGCAAAAAGAAAAGTGAGTTTTGTTTTGATCCATAGTTCTGCAGGAG
TGGTTCTTGTTTATTGTAATGCAGGCGTTTCCAGGGCTGCTGCAATTGTAATAGGTTTCTGATGAATTC
TGAACAAACCTCATTACCAAGTGCTTTTTCTTTGGTGAAAAATGCAAGACCTTCCATATGTCCAATTCT
GGCTTCATGGAGCAGCTTCGTACATATCAAGAGGGCAAAGAAAGCAATAAGTGTGACAGAATACAGGAGA
ACAGTTTCATGAGTTGCATTGTAGCAGACAATGGACAACCTGTAGTTTCTGAATTGACTTCTATAGCCATCT
TTTCCCT
```

A NOV78 polypeptide (SEQ ID NO:262) encoded by SEQ ID NO:261 is 224 amino acids in length and is presented using the one-letter amino acid code in Table 78B. The Psort profile for NOV78 predicts that this sequence has no signal peptide and is likely to be localized to the endoplasmic reticulum (membrane) with a certainty of 0.6400. In alternative embodiments, a NOV78 polypeptide is located to the plasma membrane with a certainty of 0.4960, or to the nucleus with a certainty of 0.2420.

Table 78B. NOV78 Polypeptide Sequence (SEQ ID NO:262)

MIASLHSSLGDRKRPCLCMMIIIIKREREREKTYRRRSTSGVGLKQYYHSSESHLNFLPF
LLPISTPQSLFRYPNSWDLKQSSFFFLFKVTHILNVAYGVENAFLSDFTYKISISILDLPE
TNILSYFPECFEFIEEAKRKVSFVLIHSSAGVVLVHCNAGVSRAAAIVIGFLMNSEQTSF
TSAFSLVKNARPSICPNSGFMEQLRTYQEGKESNKCDRIQENSS

A BLAST analysis of NOV78 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV78 had high
5 homology to other proteins as shown in Table 78C.

Table 78C. BLASTX results from PatP database for NOV78

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAB29109 Human cellular proliferative response protein	618	4.0e-60
patp:AAB73224 Human phosphatase AI031656_h - <i>Homo sapiens</i>	618	4.0e-60
patp:AAB73215 Murine phosphatase AA274457_m	508	1.8e-48
patp:AAM42211 Human polypeptide	290	2.3e-25
patp:AAB94018 Human protein sequence	201	6.2e-16

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 103 of 160 bases (64%) identical to a gb:GENBANK-
10 ID:HS106C24|acc:Z83313.1 mRNA from *Homo sapiens* (Human DNA sequence from PAC 106C24, between markers DXS294 and DXS730 on chromosome X). The full amino acid sequence of the protein of the invention was found to have 71 of 172 amino acid residues (41%) identical to, and 99 of 172 amino acid residues (57%) similar to, the 203 amino acid residue ptrn:SPTREMBL-ACC:Q9NGL1 protein from *Drosophila melanogaster* (Fruit fly)
15 (MAP KINASE PHOSPHATASE-1). NOV78 also has homology to the other proteins shown in the BLASTP data in Table 78D.

Table 78D. NOV78 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 18146956 d bj BAB82499.1 (AB038770)	protein phosphatase [<i>Homo sapiens</i>]	217	126/136 (92)	126/136 (92)	6e-65
gi 13277360 r ef NP_077758. 1 (NM_024438)	dual-specificity phosphatase [<i>Mus musculus</i>]	220	103/131 (78)	113/131 (85)	3e-54

gi 12845353 d bj BAB26718.1 (AK010127)	Dual specificity protein phosphatase containing protein [<i>Mus musculus</i>]	220	102/131 (77)	113/131 (85)	1e-53
gi 12858039 d bj BAB31181.1 (AK018369)	Dual specificity protein phosphatase containing protein	162	103/131 (78)	113/131 (85)	1e-52
gi 18148911 d bj BAB83499.1 (AB063187)	SKRP1 [<i>Homo sapiens</i>]	166	74/74 (100)	74/74 (100)	7e-37

This BLASTP data is displayed graphically in the ClustalW in Table 78E. A multiple sequence alignment is given, with the NOV78 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 78D.

Table 78E. ClustalW Alignment of NOV78	
NOV78	(SEQ ID NO:262)
gi 18146956	(SEQ ID NO:689)
gi 13277360	(SEQ ID NO:690)
gi 12845353	(SEQ ID NO:691)
gi 12858039	(SEQ ID NO:692)
gi 18148911	(SEQ ID NO:693)
NOV7810.....20.....30.....40.....50
gi 18146956	MIASLHSSLGDRKRPC---CMMITITKREREREKTYSRSTSGVGLKQY
gi 13277360	-MYSLNQEIKAFSRNNLRKQCTRVTTLTGKKLE-TWKDARIHVVEVEP
gi 12845353	-MHSLNQEIKAFSRDNLRKQCTRVTTLTGKKLE-TWEDATVHVVEP-EP
gi 12858039	-MHSLNQEIKAFSRDNLRKQCTRVTTLTGKKLE-TWEDATVHVVEP-EP
gi 18148911	-----VLFS-----GRGEVVTA-----
NOV7860.....70.....80.....90.....100
gi 18146956	YHSSES-HLNFIPFLPTSTPQSLFRYPNSWDLKQSSFFFLKVTILNV
gi 13277360	SSGGGCGYVQDLSSDLQGVVVKPWLLGSQDAHDLDLTKKVKVTHILNV
gi 12845353	SSGGGCGYVQDLSSDLQGVVVKPWLLGSQDAHDLELLRKHVKVTHILNV
gi 12858039	SSGGGCGYVQDLSSDLQGVVVKPWLLGSQDAHDLELLRKHVKVTHILNV
gi 18148911	G-ASSR-----SQDAHDLELLRKHVKVTHILNV
NOV78110.....120.....130.....140.....150
gi 18146956	AYGVENAFLSDFTYKTSISILDLPETNILSYFPECFEFIEBAKRKVSFVLI
gi 13277360	AYGVENAFLSDFTYKTSISILDLPETNILSYFPECFEFIEBAKRK-----
gi 12845353	AYGVENAFLSDFTYKTSISILDVPETNILSYFPECFEFIECAKDK-----
gi 12858039	AYGVENAFLSDFTYKTSISILDVPETNILSYFPECFEFIECAKDK-----
gi 18148911	-----KD-----
NOV78160.....170.....180.....190.....200
gi 18146956	HSSAGVVLVHCNAGVSRAAAIVIGFLMNSEQTSFTSAFSLVKNARPSICP
gi 13277360	----GVVLVHCNAGVSRAAAIVIGFLMNSEQTSFTSAFSLVKNARPSICP
gi 12845353	----GVVLVHCNAGVSRAAAIVIGFLMSSEETFTTALSLVKEARPSICP
gi 12858039	----GVVLVHCNAGVSRAAAIVIGFLMSSEETFTTALSLVKEARPSICP
gi 18148911	----GVVLVHCNAGVSRAAAIVIGFLMNSEQTSFTSAFSLVKNARPSICP

		210	220	230	
NOV78		NSGFMEQLR	TYQEGKES	NKCDRIQEN	SS----
gi 18146956		NSGFMEQLR	TYQEGKES	NKCDRIQEN	SS----
gi 13277360		NPGFMEQLR	TYQVGKES	NGGDKVPAED	ITGGL
gi 12845353		NPGFMEQLR	TYQVGKES	NGGDKVPAED	ITGGL
gi 12858039		NPGFMEQLR	TYQVGKES	NGGDKVPAED	ITGGL
gi 18148911		NSGFMEQLR	TYQEGKES	NKCDRIQEN	SS----

Table 78F lists the domain description from DOMAIN analysis results against NOV78. This indicates that the NOV78 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 78F. Domain Analysis of NOV78					
gnl Pfam pfam00782, DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region. SEQ ID NO:872					
CD-Length = 139 residues, 80.6% aligned Score = 112 bits (280), Expect = 2e-26					
NOV78:	88	FKVTHILNVAYGVENAF	LSDFTYKSI	SILDLPETN	ILSYFPECFEFIEEAKRKVSFVLIH 147
Sbjct:	26	LGITHVINVT	EEVPNSKNSGFLYL	HIPVDDNHETDISPYL	DEAVEFIEDARQK----- 78
NOV78:	148	SSAGVVLVHCNAGVSR	AAIVIGFLM	NSEQTSFTSAFSLV	KNARPSICPNSGFMEQLR
Sbjct:	79	--GGKVLVHCQAGIS	RSATLI	IAYLMKTRNLSL	NEAYSFVKERRPIISP
NOV78:	208	Q	208		
Sbjct:	137	E	137		

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors.

NOV78 is predicted to be expressed in at least the following tissues: parathyroid gland, peripheral blood, whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV78 is provided in Example 2.

The NOV78 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example,

brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. NOV78 nucleic acids encoding the MAP kinase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 78A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 78A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 78A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 36% of the bases may be so changed.

The novel protein of the invention includes the dual specificity phosphatase-like protein whose sequence is provided in Table 78B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 78B while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 59% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV79

The disclosed NOV79 (alternatively referred to herein as CG56862-01) includes the 939 nucleotide sequence (SEQ ID NO:263) shown in Table 79A. A NOV79 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 19-21 and ends with a stop codon at nucleotides 928-930. The disclosed NOV79 maps to human chromosome 20.

Table 79A. NOV79 Nucleotide Sequence (SEQ ID NO:263)

TACAAGCAGGGTCTCCCTATGTTTCCAGGCTGGTCTCAAACCTCCTGGGCTCACACAGTCTCTGCCTT
GGCTCCCAAAGTTCTGGGATTACAGACCCTGCAGGCGTCGGGCGTGGGCGTACGGGACGTTGTACCG
GATCGCTTCCCGGGCGGAGCTGGGGGTGCACCCGACCGCCCGGGGATCATGGGCAATGGCATG
ACCAAGGTACTTCTGGACTCTACCTCGGAAACTTCATTGGTCATCCGCCAGCCAGATTGGCTCAAGCA
TCCTGTTTCTTTCAGATGCCAAAGACCTGGATCAGCTGGGCCGAAATAAGATCACACACATCATCTCTAT
CCATGAGTCACCCAGCCTCTGCTGCAGGATATCACCTACCTTCGCATCCCGGTCGCTGATACCCCTGAG
GTACCCATGAAAAAGCACTTCAAAGAATGTATCAACTTCATCCACTGCTGCCGCTTAATGGGGGAACT
GCCTTGTGCACACCACGATTGTGACAGCGTATGTGATGACTGTGACGGGGCTAGGCTGGCGGGACGTGCT
TGAAGCCATCAAGGCCACCAGGCCCATCGCCAACCCCAACCCAGGCTTTAGGCAGCAGCTTGAAGAGTTT
GGCTGGGCCAGTTCCCGAAGGTACAGCTTCGCCGGCAGCTGGAGGAGCGCTTCGGCGAGAGCCCCCTCC
GCGACGAGGAGGAGTTGCGCGCGCTGCTGCCGCTGTGCAAGCGCTGCCGGCAGGGCTCCGCGACCTCGGC
CTCTCCGCCGGGCCGCACTCAGCAGCCTCCGAGGGAACCGTGCAGCGCTGGTGCCGCGCAGCCCCGG
GAAGCCCACCGGCCGCTGCCGCTGCTGGCGCGCTCAAGCAGACTTCTCTTGCTCCCCCGGTGTCTGT
CCCGCAAGGGCGGCAAGTGAGGATGCAGT

A NOV79 polypeptide (SEQ ID NO:264) encoded by SEQ ID NO:263 is 303 amino acids in length and is presented using the one-letter amino acid code in Table 79B. The Psort profile for NOV79 predicts that this sequence has a signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500. In alternative embodiments, a NOV79 polypeptide is located to lysosomes with a certainty of 0.2216. The Signal P predicts a likely cleavage site for a NOV79 peptide is between positions 24 and 25, *i.e.*, at the dash in the sequence VLG-LQ.

Table 79B. NOV79 Polypeptide Sequence (SEQ ID NO:264)

MFRLVSNWAHTVLLPWPPKVLGLQTLQASGLGRQSCDRIASRAASWGCTRTAAPGIM
GNGMTKVLPGLYLGNFIGHPASQIGSSILFLSDAKDLQDLGRNKITHIISIHESQPPLQ
DITYLRI PVADTPEVPMKKHFKECINFHCCRLNGGNCLVHTTIVTAYVMTVTGLGWRDV
LEAIKATRPANPNPGFRQQLLEFGWASSQKVQLRRQLEERFGESPFRDEEELRALLPLC
KRCRQSATSASSAGPHSAASEGTVQRLVPRTPREAHRLPLLARVKQTFSLPRCLSRK
GGK

A BLAST analysis of NOV79 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV79 had high homology to other proteins as shown in Table 79C.

Table 79C. BLASTX results from PatP database for NOV79

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAE04840 Human SGP008 phosphatase polypeptide	1298	3.5e-132
patp:AAY68795 Amino acid sequence of a human protein	433	1.6e-40
patp:AAB67167 Human dual-specificity phosphatase DSP-3	433	1.6e-40
patp:AAB66431 Human DSP-3 protein - <i>Homo sapiens</i> , 184 aa.	433	1.6e-40
patp:AAB73216 Human phosphatase AA374753_h - <i>Homo sapiens</i>	433	1.6e-40

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 93 of 118 bases (78%) identical to a gb:GENBANK-ID:HUMFLNG6PD|acc:L44140.1 mRNA from *Homo sapiens* (chromosome X region from filamin (FLN) gene to glucose-6-phosphate dehydrogenase (G6PD). The full amino acid sequence of the protein of the invention was found to have 273 of 276 amino acid residues (98%) identical to, and 274 of 276 amino acid residues (99%) similar to, the 275 amino acid residue ptnr:TREMBLNEW-ACC:CAC10008 protein from *Homo sapiens* (Human) (BA243J16.6 (NOVEL PROTEIN)). NOV79 also has homology to the other proteins shown in the BLASTP data in Table 79D.

Table 79D. NOV79 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 18104942 ref NP_542178.1 (NM_080611)	dual specificity phosphatase-like 15 [<i>Homo sapiens</i>]	243	241/245 (98)	242/245 (98)	e-133
gi 17458347 ref XP_059288.1 (XM_059288)	similar to BA243J16.6 (novel protein with a dual specificity phosphatase, catalytic domain) (<i>H. sapiens</i>) [<i>Homo sapiens</i>]	235	226/252 (89)	227/252 (89)	e-121
gi 9910432 ref NP_064570.1 (NM_020185)	mitogen-activated protein kinase phosphatase x; homolog of mouse dual specificity phosphatase LMW-DSP2; JNK-stimulating phosphatase 1 [<i>Homo sapiens</i>]	184	88/186 (47)	119/186 (63)	3e-44

gi 13183069 gb AAK15038.1 AF237619_1 (AF237619)	dual specificity phosphatase TS-DSP2 [<i>Mus musculus</i>]	184	87/186 (46)	119/186 (63)	1e-43
gi 14726046 ref XP_046543.1 (XM_046543)	mitogen-activated protein kinase phosphatase x [<i>Homo sapiens</i>]	184	86/186 (46)	112/186 (59)	2e-41

This BLASTP data is displayed graphically in the ClustalW in Table 79E. A multiple sequence alignment is given, with the NOV79 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 79D.

Table 79E. ClustalW Alignment of NOV79

NOV79	(SEQ ID NO:264)
gi 18104942	(SEQ ID NO:694)
gi 17458347	(SEQ ID NO:695)
gi 9910432	(SEQ ID NO:696)
gi 13183069	(SEQ ID NO:697)
gi 14726046	(SEQ ID NO:698)

1020304050

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV79

MFPRLVSNWAHTVLLPWPPKVLGLQTLQASGLGRQGSCDRIASRAASWG

gi|18104942|

gi|17458347|

gi|9910432|

gi|13183069|

gi|14726046|

60708090100

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV79

CTRTAAPGIMGNMGMTKVLPGLYLGNFIGHHPASQIGSSILFLSDAKDLDQL

gi|18104942|

-----MGNGMTKVLPGLYLGNFIGHHPASQIGSSILFLSDAKDLDQL

gi|17458347|

-----MGNGMTKVLPGLYLGNFI-----DAKDLDQL

gi|9910432|

-----MGNGMNTKILPGLYLGNFK-----DARDAEQL

gi|13183069|

-----MGSGMSQILPGLYLGNFK-----DARDAEQL

gi|14726046|

-----MGNGMNTKILPGLYLGNFK-----DARDAGQL

110120130140150

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV79

GRNKHITHISIHESPOELLQDITYLRIPVADTPEVPMKKHFKECINFIHC

gi|18104942|

GRNKHITHISIHESPOELLQDITYLRIPVADTPEVPMKKHFKECINFIHC

gi|17458347|

GRNKHITHISIHESPOELLQDITYLRIPVADTPEVPMKKHFKECINFIHC

gi|9910432|

SKNKVTHILSVHDSARPMLEGVKYLCPAADSPPSONLTRHFKESIKFIHE

gi|13183069|

SRNKVTHILSVHDTAREMLEGVKYLCPAADTPSONLTRHFKESIKFIHE

gi|14726046|

SRNKVTHILSVHDXXXXXXXGVKYLCPAADSPPSONLTRHFKESIKFIHE

160170180190200

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV79

CRLNGGNCLVH-----TTHIVTAYVMTVTGLGWRDVLEAIKATRPAN

gi|18104942|

CRLNGGNCLVHC-----TTHIVTAYVMTVTGLGWRDVLEAIKATRPAN

gi|17458347|

CRLNGGNCLVHCFAGLSRSTTTHIVTAYVMTVTGLGWRDVLEAIKATRPAN

gi|9910432|

CRLRGESCLVHCLAGVSRSVTLVIAYIMTVDTFGWEDALHTVRAGRSCAN

gi|13183069|

CRLQGESCLVHCLAGVSRSVTLVIAYIMTVDTFGWEDALHTVRAGRSCAN

gi|14726046|

CRLRGESCLVHCLAGVSRSVTLVIAYIMTVDTFGWEDALHTVRAGRSCAN

210220230240250

NOV79	PNPGFRQOLEEFGWASSQVQLRRLLEERFGESPFRRDEEELRALLPLCKR
gi 18104942	PNPGFRQOLEEFGWASSQ--KLRLLEERFGESPFRRDEEELRALLPLCKR
gi 17458347	PNPGFRQOLEEFGWASSQ--KLRLLEERFGESPFRRDEEELRALLPLCKR
gi 9910432	PNVGFQRLQLEFEKHEVH--CYRQWLKEEYGESPLQDAEEAKNIL----
gi 13183069	PNLGFQRLQLEFEKHEVH--CYRQWLKEEYGENPLRDAEEAKNIL----
gi 14726046	PNVGFQRLQLEFEKHEVH--CYRQWLKEEYGESPLQDAEEAKNIL----
	260 270 280 290 300
NOV79	CRQGSATSASSAGPHSAASECTVQRLVPRTPREAHRPLPLLARVKQTESC
gi 18104942	CRQGSATSASSAGPHSAASECTVQRLVPRTPREAHRPLPLLARVKQTESC
gi 17458347	CRQGSATSASSAGPHSAASECTVQRLVPRTPREAHRPLPLLARVKQTESC
gi 9910432	-----AAPCILK-----EWA
gi 13183069	-----AAPCILK-----EWA
gi 14726046	-----AAPCIMK-----EWA
	310
NOV79	LPRCLSRKGGK
gi 18104942	LPRCLSRKGGK
gi 17458347	LPRCLSRKGGK
gi 9910432	FLRRL-----
gi 13183069	FLRRL-----
gi 14726046	FLRRL-----

Table 79F lists the domain description from DOMAIN analysis results against NOV79. This indicates that the NOV79 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 79F. Domain Analysis of NOV79				
gnl	Smart	smart00195, DSPc, Dual specificity phosphatase, catalytic domain		
SEQ ID NO:873				
CD-Length = 139 residues, 97.8% aligned				
Score = 108 bits (271), Expect = 3e-25				
NOV79:	63	GMTKVLPGLYLGNFIGHPASQIGSSILFLSDAKDLQGRNKITHIIS-IHESQPPLLQD		121
Sbjct:	1	G +++LP LYLG++ SDA +L L + ITH+I+ E P		45
NOV79:	122	ITYLRIPVADTPEVPMKKHFKECINFHCCRLNGGNCVLH-----TTIVTAYVMTVT		173
Sbjct:	46	FLYLGI PVDDNTETKISPYLPEAVEFIEDAEKKGGKVLVHCQAGVSRSATLIIAYLMKYR		105
NOV79:	174	GLGWRDVL EAIKATRP IANPNPGFRQOLEEF		204
Sbjct:	106	+ D + +K RPI +PN GF +QL E+		136

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors.

10

NOV79 is predicted to be expressed in at least the following tissues: brain, kidney, pancreas, testis, whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV79 is provided in Example 2.

The NOV79 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. The NOV79 nucleic acid encoding the phosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 79A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 79A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 79A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 22% of the bases may be so changed.

The novel protein of the invention includes the dual specificity phosphatase-like protein whose sequence is provided in Table 79B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 79B while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 2% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV80

The disclosed NOV80 (alternatively referred to herein as CG56882-01) includes the 2039 nucleotide sequence (SEQ ID NO:265) shown in Table 80A. A NOV80 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 100-102 and ends with a TGA codon at nucleotides 1947-1949. The disclosed NOV80 maps to human chromosome 10.

Table 80A. NOV80 Nucleotide Sequence (SEQ ID NO:265)

```
CATGTGCTAGGTTATTCCAGTGCAGAGGCCACACTTGGGCCGTGCGAGCAGCCCTCCTCACTTCAGGGG
TCACCCCTCCCCAAGACCCATTGCCCCATCATGGCCGGGACCGGCTCCCCAGGAAGGTGATGGACGCCAA
GAAGCTGGCCAGCCTGCTGCGGGGCGGGCCTGGGGGGGGCCTGGTCATCGACAGTCACTCCTTCTGGAG
TACAACAGCTGGCATGTGCTCAGCTCCGTCAACATCTGCTGCTCCAAGCTGGTGAAGTGGCGGTTGCAGA
AGGGCAAGGTGACCATTTGTGGAGTTTCATCCAGCCGGCCGACGCAGCCAGGTGGAGGCCACTGAGCCACA
GGACGTGGTGGTCTATGACCAGAGCACGCGGGCCGCGAGACAGCTTCCTCTCCATCCTGCTGAGCAAGCTG
GATGGCTGCTTCCACAGCGTGGCCGGCTGCTTCCACAGCATGGCCATCATCAGGGGGGCTTCGCCACCT
TCTCCTCCTGCTTCCCGACCTTGCGAGGGCGAGCCTGCTGCCCTGCTACCCATGAGCCTCTCCAGTC
CTGCCTGCTCGTGCCAGCGTGGGCCCTGACCTCATCCTGCCTCACCTCTACCTGGGCTCGCAGGAAGAC
GTCTGAACAAGGATCTGATGACGCAGAATGGAATAAGCTACGTCTCTATGCCAGCAACTCCTGCCCA
AGCCTGACTTCATCTACCAGAGCCACTTCTTGCGGGTCCCATCAACGACAACCTACTGTGAAAAGCTGCT
GCCCTGGCTGGACAAGTCCATCGAGTTCTGTCGATAAAGCCAAGCTGTCCAGCTGCCAAGTCATCGTCCAC
CGTCTGGCCGGCATCTCTGCTGTGCCACTATCGCCATCGCCTACATCATGAAGACCATGGGCATGTCTCT
CCGAAGACGCTACAGGTTTGTGAAGGACCAGCGCCCGTCCATCTCGCCCACTTCAACTTCTGGGCCA
GCTGCTGGAGGACCAGAGCAGCCGAAGCTGCTGGCCGCGGTGACAGGGCGAGCGGGCACCCCTCAGGA
ATGCAGGAGCTCCCCCAGCCCTGCGGCCGGGGCCCCACTGCCATGGCTGCCACCACCTACCTCAGAGA
CCGTGCCACCAGGAGTGCGGCTGCCAGGAGGGCGGGCCGAGCGGGCGAGGAAGCCCGGGCGCCCCC
CACGGCCACCAGCAGCTGCAGCAGGGCCTGCGCAGCCTGCGCCTCTCTCGGACCACCTGCAGGACACC
AGCGCCTCAAGCCCTCCTTCTCTCTGACATCAAGTGGCCTACGCCCCAGCAGGCGGGCCCGGGCGCC
CGGGCCAGCGACCCCGGCGAGGCCCGAAGCTCTCTGAAAGCTGGACAGCCAGTCCGGGCCATGTTGGG
GTGCTGTTTGGCCCTTTCGGCCGGGCGGGCGCCCCGGAACCCAAACGGCTGCAGCGACCTGCCACGGCGGG
AGGCAGCAAGGCTGAGCCCGGACGGGTGACGAGAGCTGGCCGACGAGCTGGCCCCGATTTCGCACTT
CAAGTGCTGCAGCTGCCAGATGGAGTTCGAGGAGGGCATGGTGGAGGGGCGCGCGCGGCGAGGAGCTG
GCCGCCCTGGGCAAGCAGGGGAGCTTCTCGGGCAGCGTGGAGGTTCAGAGATGTCGAGATGTCGACCCCTCCGT
GCCCTCGGCTCCGCCGCCGACGCTGGGCAGTTATAATATATATATATATATATGCAAAGAAAGGCCAA
TGGTTTAC
```

A NOV80 polypeptide (SEQ ID NO:266) encoded by SEQ ID NO:265 is 616 amino acids in length and is presented using the one-letter amino acid code in Table 80B. The Psort profile for NOV80 predicts that this sequence is a Type Ib membrane protein, has no signal peptide, and is likely to be localized at the plasma membrane with a certainty of 0.7000. In alternative embodiments, a NOV80 polypeptide is located to peroxisomal microbodies with a certainty of 0.3000, to the mitochondrial inner membrane with a certainty of 0.2143, or to the nucleus with a certainty of 0.3000.

Table 80B. NOV80 Polypeptide Sequence (SEQ ID NO:266)

MAGDRLPRKVM DAKKLASLLRGPGGGLVIDSHSFLEYNSWHVLSSVNICCSKLVKWR LQ
KGKVTIVEFIQPAARSQVEATEPQDVVVDQSTRAADSFLSILLSKLDGCFHSVAGCFHS
MAIITGGFATFSSCFDLCEGEPAALLPMSLSQSCLLVPSVGLTLILPHLYLG SQEDVLN
KDLMTQNGISYVLYASNSCPKPDFIYQSHFLRVPINDNYCEKLLPWLDKSI EFVDKAKLS
SCQVIVHRLAGISCCATIAIAYIMKTMGMSSDAYRFVKDQRPISPNFNFLGQLLEDQS
SPKLLAAVQGDAGTPSGMQEPPPSPAAGAPLPWLPPPTSETAATRSAAAREGGPSAGRKP
PAPPTATSTLQQGLRSLRLSSDHLQDTSRLKPSFSLDIKSAYAPSRPPGGPGPATPARPR
SSLKAGQPVGAMLG LPSPCPDAAAPAARQPRRGPEASQSGPCWAPARTPRPGTPTARRA
TPRAALTSATRLPGPGQPASPGAWTPPLDSLKRPRCFSP EGVQGPGRVLFAPFGRAGAPE
PNGCSDLPRREAARAEPTGTQTSWPELAPDSHFKCCSCQMEFEEGMVEGRARGEELAL
GKQGSFSGSVEVIEMS

A BLAST analysis of NOV80 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV80 had high homology to other proteins as shown in Table 80C.

Table 80C. BLASTX results from PatP database for NOV80

		High Score	Smallest Sum Probability P (N)
Sequences producing High-scoring Segment Pairs:			
patp:AAW29150	Dual-specific murine thr-tyr phosphatase	1873	4.1e-193
patp:AAE04834	Human SGP002 phosphatase polypeptide	950	6.6e-108
patp:AAU09016	Human dual specificity phosphatase 21117	950	6.6e-108
patp:AAM25744	Human protein sequence	955	7.8e-96
patp:AAB20325	Human protein phosphatase and kinase protein	949	3.4e-95

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1587 of 1930 bases (82%) identical to a gb:GENBANK-ID:HSU27193|acc:U27193.1 mRNA from *Homo sapiens* (Human protein-tyrosine phosphatase mRNA). The full amino acid sequence of the protein of the invention was found to have 489 of 625 amino acid residues (78%) identical to, and 514 of 625 amino acid residues (82%) similar to, the 625 amino acid residue ptnr:SWISSNEW-ACC:Q13202 protein from


```

gi|4758212|-----MAGDRLPRKVM DAKK L A S L L R G G P G
gi|6679156|-----MAGDRLPRKVM DAKK L A S L L R G G P G
gi|17471343|-----MAGDRLPRKVM DAKK L A S L L R G G P G
gi|13639013|-----
gi|12697945|AFEHTTITSSWQIKEGGK RGLIVVMAHEMGTQIVTER-LVALL E S C T E

      60      70      80      90      100
NOV80      G G L V I D S H S F E Y N S W H V L S S V N I C C S K L V K R L O K G K V T I V E F T I Q P A A R
gi|4758212| G G L V I D S R S F E Y N S W H V L S S V N I C C S K L V K R R L Q O G K V T I A E L I Q P A A R
gi|6679156| G G L V I D S R S F E Y N S W H V L S S V N I C C S K L V K R R L Q O G K V T I A E L I Q P A T R
gi|17471343| G-----A W S S T V T P-----S W S T T A G-----
gi|13639013|-----
gi|12697945|K V L L I D S R P F E Y N T I S H I L E A I N I N C S K L M K R R L Q O D K V L I T E L I Q H S A K

      110     120     130     140     150
NOV80      S Q V E A T E P Q D V V V Y D Q S T R A A D S F L S I L L S K L D G C F H S V A G C F H S M A I I T
gi|4758212| S Q V E A T E P Q D V V V Y D Q S T R D A S V L A A D--S F L S I L L S K L D G C F D S V A I I T
gi|6679156| S Q V E A T E P Q D V V V Y D Q S T R D A S V L A A D--S F L S I L L S K L D G C F D S V A I I T
gi|17471343| S-----M C S A P-----S T S A A P-----S-----C V A G C F H S M A I I T
gi|13639013|-----
gi|12697945|H K V I D I D C S Q K V V V Y D Q S S Q D V A S L S S D--C F L T V L L G K L E K S F N S V H L L A

      160     170     180     190     200
NOV80      G G F A T F S S C F P D L C E G E P A A L L P M S L S Q S C L L V P S V G L T L I L P H L Y L G S Q
gi|4758212| G G F A T F S S C F P L C E G K P A A L L P M S L S Q P C L P V P S V G L T R I L P H L Y L G S Q
gi|6679156| G G F A T F S S C F P L C E G K P A T L P S M S L S Q P C L P V P S V G L T R I L P H L Y L G S Q
gi|17471343| G G F A T F S S C F P D L C E G E P A A L L P M S L S Q S C L L V P S V G L T L I L P H L Y L G S Q
gi|13639013|-----M S L S Q P C L P V P S V G L T R I L P H L Y L G S Q
gi|12697945| G G F A E S R C F P L C E G K--S T I V E T C I S Q P C L P V A N I G P T R I L P N L Y L G C Q

      210     220     230     240     250
NOV80      E D V L N K D L M T Q N G I S Y V L Y A S N S C P K P D F I Y Q S H E L R V P I N D N Y C E K L L P
gi|4758212| K D V L N K D L M T Q N G I S Y V L N A S N S C P K P D F I C E S R F M R V P I N D N Y C E K L L P
gi|6679156| K D V L N K D L M T Q N G I S Y V L N A S N S C P K P D F I C E S R F M R I P I N D N Y C E K L L P
gi|17471343| E D V L N K D L M T Q N G I S Y V L Y A S N S C P K P D F I Y Q S H E L R V P I N D N Y C E K L L P
gi|13639013| K D V L N K D L M T Q N G I S Y V L N A S N S C P K P D F I C E S R F M R V P I N D N Y C E K L L P
gi|12697945| R D V L N K E L M Q O N G I G Y V L N A S N I C P K P D F I P E S H E L R V P V N D S F C E K I L P

      260     270     280     290     300
NOV80      W L D K S I E F V D K A L S S C Q V I V H R L A G I S C C A T I A I A Y I M K T M G M S S D A Y
gi|4758212| W L D K S I E F I D K A L S S C Q V I V H C L A G I S R S A T I A I A Y I M K T M G M S S D D A Y
gi|6679156| W L D K S I E F I D K A L S S C Q V I V H C L A G I S R S A T I A I A Y I M K T M G M S S D D A Y
gi|17471343| W L D K S I E F V D K A L S S C Q V I V H R L A G I S C C A T I A I A Y I M K T M G M S S D A Y
gi|13639013| W L D K S I E F I D K A L S S C Q V I V H C L A G I S R S A T I A I A Y I M K T M G M S S D D A Y
gi|12697945| W L D K S V D F I E K A K A S N G C V I V H C L A G I S R S A T I A I A Y I M K R M D M S L D E A Y

      310     320     330     340     350
NOV80      R F V K D R P S I S P N F N F L G Q L L E D Q S S P K L L A A V O G D A G-----T P S G M Q E
gi|4758212| R F V K D R P S I S P N F N F L G Q L L E Y E R T L K L L A A L Q G D P G-----T P S G T P E
gi|6679156| R F V K D R P S I S P N F N F L G Q L L E Y E R S E K L L A A L Q T D-G-----P H L G T P E
gi|17471343| R F V K D R P S I S P N F N F L G Q L L E D Q S S P K L L A A V O G D A G-----T P S G M Q E
gi|13639013| R F V K D R P S I S P N F N F L G Q L L E Y E R S E K L L A A L Q G D P G-----T P S G T P E
gi|12697945| R F V K E K R P T I S P N F N F L G Q L L D Y E K K I K N O T G A S C P K S K L K L L H L E K P N E

      360     370     380     390     400
NOV80      P P P S P A A G--A P L E W L P P P T S E T A A T R--S A A A R E G G P--S A G R K P P A P P
gi|4758212| P P P S P A A G--A P L E R L P P P T S E S A A T G--N A A A R E G G L--S A G G E P P A P P
gi|6679156| P L M G P A A G--I P L P R L P P S T S E S A A T G S E A A T A A R E G S P--S A G G D A P I P S
gi|17471343| P P P S P A A G--A P L E W L P P P T S E T A A T R--S A A A R E G G P--S A G R K P P A P P
gi|13639013| P P P S P A A G--A P L P R L P P P T S E S A A T G--N A A A R E G G L--S A G G E P P A P P
gi|12697945| P V P A V S E G G Q K S E T P L S P P C A L S A T S E--A A G O R P V H P A S V P S V P S V Q P

```


	410	420	430	440	450
NOV80	T--ATSTLQQGLRSLRLSSDHLQDTSRLKPSFSLDIKSA-YAPSRPRGGP				
gi 4758212	TPPATSAALQQGLRGLHLSSDRLQDTNRLKRSFSLDIKSA-YAPSRPRDGP				
gi 6679156	TAPATSALQQGLRGLHLSSDRLQDTNRLKRSFSLDIKSA-YAPSRPRDGP				
gi 17471343	T--ATSTLQQGLRSLRLSSDHLQDTSRLKPSFSLDIN-----				
gi 13639013	TPPATSAALQQGLRGLHLSSDRLQDTNRLKRSFSLDIKSA-YAPSMRPRDGP				
gi 12697945	SLLEDSPILVQALSGHLSDRLQDTSRLKRSFSLDIKSVSYASMAASLH				

	460	470	480	490	500
NOV80	GPATPARPRSSLRAGQPVGAM--LGLSPSPCPDAAPAARACRPRRGPE--A				
gi 4758212	GPPDPGEAPKLCGLDSPSGAA--LGLSPSPSPDPAAPAEARPRRRRRPR				
gi 6679156	GPPDPGEAPKLCGLDSPSGGT--LGLSPSPSPDPSVPECRPRRRR--RP				
gi 17471343	-----				
gi 13639013	GPPDPGEAPKLCGLDSPSGAA--LGLSPSPSPDPAAPAEARPRRRRRPR				
gi 12697945	GFSSSEDALEYKPSSTTLTGNTKLCQSPVQELSEOTPETSEDKEEASIP				

	510	520	530	540	550
NOV80	SQSGPCW--APARTPRPGTPTARRATPRAAL--TSATR-----				
gi 4758212	PAGSPAR--SPAHSGLGNFGDAAROTPRHGLSALSAPG-----				
gi 6679156	PASSPAR--SPAHSGLGNFGDTAROTPRHGLSALSAPG-----				
gi 17471343	-----				
gi 13639013	PAGSPAR--SPAHSGLGNFGDAAROTPRHGLSALSAPG-----				
gi 12697945	KKLQTARPSDSQSKRLHSVRTSSSGTAQRSLSPHRSQSVEDNYHTSFL				

	560	570	580	590	600
NOV80	--LPGPGQPASPGA-----WTPPLDSLKR--PR-----CFSPEGVQCPG				
gi 4758212	--LPGPGQPACPGA-----WAPPLDSPGT--PSPDGPWCFSPEGAQCAG				
gi 6679156	--LPGPGQPACPCG-----WVPLDSPGT--PSPDGPWCFSPEGAQCAG				
gi 17471343	-----				
gi 13639013	--LPGPGQPACPGA-----WAPPLDSPGT--PSPDGPWCFSPEGAQCAG				
gi 12697945	FGLSTSQOHLTKSAGLGLKGWHSDDLAFQTSTPSLTSSWYFATESSHFYS				

	610	620	630	640	650
NOV80	R-VLFAPFGR--AGAPEPN-----GCSDLPRREA				
gi 4758212	G-VLFAPFGR--AGAPGPG-----GCSDLRRREA				
gi 6679156	--AVFSAFGRVSAGAPGPGNSSSSGGGGGGGGGGGGGGGSSSSSSSS				
gi 17471343	R-VLFAPFGR--AGAPEPN-----GCSDLPRREA				
gi 13639013	G-VLFAPFGR--AGAPGPG-----GCSDLRRREA				
gi 12697945	ASATYGGASASYAYSQSLP-----TCGDQVYSVR				

	660	670	680	690	700
NOV80	ARAEPGT-----GQTSWPEDELAP-----DSHFKCCSCQ				
gi 4758212	ARAEP-----RDARTGWPEEPAP-----ETQFKRRSCQ				
gi 6679156	SSSSSSSSSSSSSSSSDLRRRDVRTGWPEEPAA-----DAQFKRRSCQ				
gi 17471343	ARAEPGTTRGP-----AELPDWQGGTTAGEELGLRGSGRRGAWLACAGSTG				
gi 13639013	ARAEP-----RDARTGWPEEPAP-----ETQFKRRSCQ				
gi 12697945	RRQKESDR-----ADSRRSWHEESPF-----EKQFKRRSCQ				

	710	720	730	740	750
NOV80	---MEFEEGMV--EGRAR--GEELAALGKQGSFSGSV-----				
gi 4758212	---MEFEEGMV--EGRAR--GEELAALGKQASFGSV-----				
gi 6679156	---MEFEEGMV--EGRAR--GEELAALGKQTSFGSV-----				
gi 17471343	RGALDCQDNEL--EKITRFTMELAKKCFIAFPFPWT-----				
gi 13639013	---MEFEEGMV--EGRAR--GEELAALGKQALXSGSVGGHRGVLTPLSSA				
gi 12697945	---MEFGESIMSENRRS--EELGKVGQSFSFGSM-----				

	760
NOV80	-----EVIEMS--
gi 4758212	-----EVIEMS--
gi 6679156	-----EVIEMS--
gi 17471343	-----NAYEIKL-

gi 13639013	PPPAAREPVINVYYI
gi 12697945	-----EIEVS--

Table 80F lists the domain description from DOMAIN analysis results against NOV80. This indicates that the NOV80 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 80F. Domain Analysis of NOV80					
gnl Smart smart00195, DSPc, Dual specificity phosphatase, catalytic domain					
SEQ ID NO:873					
CD-Length = 139 residues, 97.1% aligned					
Score = 154 bits (388), Expect = 2e-38					
NOV80:	162	GLTLILPHLYLGSQEDVLNKDLMTQNGISYVLYASNSCPKPDFIYQSHFLRVPINDNYCE	221		
Sbjct:	1	G + ILPHLYLGS D N L+ + GI++V+ + P +L +P++DN	59		
NOV80:	222	KLLPWLDKSIEFVDKAKLSSCQVIVHRLAGISCCATIAIAYIMKTMGMSSEDAYRFVKDQ	281		
Sbjct:	60	K+ P+L +++EF++ A+ +V+VH AG+S AT+ IAY+MK MS DAY FVK++	119		
NOV80:	282	RPSISPNFNFLGQLE	297		
		RP ISPNF FL QL+E			
Sbjct:	120	RPIISPNFGFLRQLIE	135		

Mitogen-activated protein (MAP) kinase phosphatases constitute a growing family of dual specificity phosphatases thought to play a role in the dephosphorylation and inactivation of MAP kinases and are therefore likely to be important in the regulation of diverse cellular processes such as proliferation, differentiation, and apoptosis. For this reason it has been suggested that MAP kinase phosphatases may be tumor suppressors.

NOV80 is predicted to be expressed in at least the following tissues: kidney. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV80 is provided in Example 2.

The NOV80 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, brain disorders including epilepsy, eating disorders, schizophrenia, ADD, and cancer; heart disease; blood disorders, kidney disorders, liver diseases, inflammation and autoimmune disorders including Crohn's disease, IBD, allergies, rheumatoid and osteoarthritis, inflammatory skin disorders, allergies, blood disorders; psoriasis; colon-, ovarian-, testicular-, lymphatic-, brain-, and pancreatic cancers; leukemia AIDS; thalamus disorders; metabolic disorders including diabetes and obesity; lung diseases such as asthma, emphysema, cystic

fibrosis, and cancer; pancreatic disorders including pancreatic insufficiency; and prostate disorders including prostate cancer and other diseases, disorders and conditions of the like. The NOV80 nucleic acid encoding the phosphatase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a dual specificity phosphatase-like protein includes the nucleic acid whose sequence is provided in Table 80A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 80A while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 80A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 18% of the bases may be so changed.

The novel protein of the invention includes the dual specificity phosphatase-like protein whose sequence is provided in Table 80B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 80B while still encoding a protein that maintains its dual specificity phosphatase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 22% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV81

NOV81 includes two galactosyltransferase-like proteins, designated herein as NOV81a and NOV81b.

NOV81a

The disclosed NOV81a (alternatively referred to herein as CG56283-01) includes the 1247 nucleotide sequence (SEQ ID NO:267) shown in Table 81A. A NOV81a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 33-35 and ends with a stop codon at nucleotides 1224-1226. The disclosed NOV81a maps to human chromosome 19.

Table 81A. NOV81a Nucleotide Sequence (SEQ ID NO:267)

```
CTCCCGCGGCGGCCCTTCCCTGGGCGGGCTCATGCGCTGCCCCAAGTGCCTTCTCTGCTGTGTCAGCACT
GCTCACACTCCTGGGCTCAAAGTGACATCGAATGGACATCCGAGTCCCGGCTCAGCAAGGCCTACCCC
AGCCCTCGGGGACCCCGCCAAGCCCCACGCCAGCCAACCTGAGCCACCTACCTGCCAACCTCTCCA
CCCGCCTGGGCGAGACTATCCCGCTGCCCTTTGCTTACTGGAACAGCAGCAGTGGCGGCTGGGGTCCCT
GCCCAGTGGGGACAGCACTGAAACGGGGGGCTGCCAGGCTTGGGGGGCCGCCGCCGCCAGAGATCCCT
GACTTCGCCTCCTACCCCAAGGACCTCCGCGCTTCTTGCTGTGTCAGCAGCCTGCCGGAGCTTCCCACAGT
GGCTGCTGGAGGTGGTGGCGGCCAAGTCTCCAGCTGCTCAGATACTGATGTCCCTACCTGCTGTGGC
CGTCAAGTCAGAACCAGGGCGCTTTCAGAACGACAGGCGCTGAGAGAGACGTGGGGCAGTCCAGTCCA
GGGATCCGGCTGCTCTTCTGCTAGGGTCTCCGGTAGGTGAGGCGGGGCTGACCTAGACTCACTAGTGG
CATGGGAGAGCCGTCGCTACAGTGACCTGCTGCTCTGGGACTTCTCGACGTCCCATTAACCAGACGCT
CAAAGACCTGCTGCTGCTGGCCTGGCTGGGCGGCCACTGCCCCACCGTGAGTTTGTCTGCGAGCTCAG
GACGATGCCTTTGTACACACCCCTGCCCTGCTGGCTCACCTGCGGGCCCTGCCACCTGCCTCGGCCGAA
GCCTCTACCTGGGTGAGGTCTTTACCCAGGCCATGCCTCTCCGGAAGCCAGGAGGACCTTCTATGTGCC
CGAGTCCTTCTTCGAAGGTGGCTACCCAGCCTATGCAAGCGGGGGTGGCTACGTATTGCCGGGCGCTG
GCACCTGGCTGCTGCGGGCGGCGAGCCCGTGTGGCACCCCTTCCCTTTGAGGACGTCTACACTGGCCTTT
GCATCCGAGCCCTGGGCTGTTGCCCGAGGCCACCCAGGCTTCTCACAGCTGGCCAGCAGACCGCAC
TGCGGACCACTGTGCTTTCGCAACCTGCTGCTGGTACGGCCCTGGGCCCCAGGCCAGCATTCCGGCTC
TGGAAACAACCTGCAAGACCCCAAGGCTCCAGTGCTGACTCTCATTGGGGAGGGCGGAG
```

A NOV81a polypeptide (SEQ ID NO:268) encoded by SEQ ID NO:267 is 397 amino acids in length and is presented using the one-letter amino acid code in Table 81B. The Psort profile for NOV81a predicts that this sequence has a signal peptide and is likely to be localized to lysosomes with a certainty of 0.8650, or to the outside of the cell with a certainty of 0.8191. The Signal P predicts a likely cleavage site for a NOV81a peptide is between positions 34 and 35, *i.e.*, at the dash in the sequence SKA-YP.

Table 81B. NOV81a Polypeptide Sequence (SEQ ID NO:268)

```
MRCPKCLLCLLSALLTLGLKVYIEWTSESRLSKAYPSPRGTPPPSPTFANPEPTLPANLST
RLGQTIPLPFAYWNQQWRLGSLPSGDSTETGGCQAWGAAAATEIPDFASYPKDLRRFLL
SAACRSFPQWLPGGGGGQVSSCSDTDVPYLLLA VKSEPGRFAERQAVRETWGSFAPGIRL
LFLLGSPVGEAGPDLSLVAVESRRYSDDLWDFLDVPFNQTLKDLLLAWLGRHCPVS
FVLRAQDDAFVHTPALLAHLRALPPASARSLYLGEVFTQAMPLRKPGGPFYVPESFFEGG
YPAYASGGGYVIAGRLAPWLLRAARVAPFPFEDVYTGLCIRALGLVPQAHGFLTAWPA
DRTADHCAFRNLLLVRPLGPQASIRLWKQLQDPRQC
```

NOV81b

The disclosed NOV81b (alternatively referred to herein as CG56283-02) includes the 1368 nucleotide sequence (SEQ ID NO:269) shown in Table 81C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 4-6 and ends with a TGA codon at nucleotides 1195-1197. In Table 81C, putative untranslated regions are indicated by underlining, and the stop and start codons are indicated in bold. The disclosed NOV81b maps to human chromosomes 19.

Table 81C. NOV81b Nucleotide Sequence (SEQ ID NO:269)

GTCATGCGCTGCCCCAAGTGCCTTCTCTGCCTGTCAGCACTGCTCACACTCCTGGGCCTC
AAAGTGATCGAGTGGACATCCGAGTCCCGGCTCAGCAAGGCTACCCAGCCCTCGG
GGCACCCTGGCCAGCTATCCCGCTGCCCTTGCTTACTGGAACAGCAGCAGTGGCGG
CTGGGGTCCCTGCCAGTGGGGACAGCACTGAAACGGGGGGCTGCCAGGCTTGGGGGGC
GCCGCCGCCACCGAGATCCCTGACTTCGCCCTCCACCCCAAGGACCTCCGCCGCTTCTTG
CTGTACAGCAGCCTGCCGGAGCTTCCACAGTGGCTGCCTGGAGGTGGTGGCAGCCAAGTC
TCCAGCTGCTCAGATACTGATGTCCCTACCTGCTGTTGGCCGTCAGTCAGAACAGGG
CGCTTTCAGAACGACAGGCCGTGAGAGAGACGTGGGGCAGTCCAGCTCCAGGGATCCGG
CTGCTCTTCCTGCTAGGGTCTCCGGTGGGTGAGGCGGGGCTGACCTAGACTCACTAGTG
GCCTGGGAGAGCCGTCGCTACAGTGACCTGCTGCTCTGGGACTTCCTCGACGTCCCATT
AACCAGACGCTCAAAGACCTGCTGCTGCTGGCTGGTGGGCCGCCACTGCCACCCGTG
AGTTTGTCTTGCAGCTCAGGACGATGCCTTGTACACACCCCTGCCCTGCTGGCTCAC
CTGCGGGCCCTGCCACTGCCTCGGCCCGAAGCCTCTACCTGGGTGAGGTCTTTACCCAG
GCCATGCCTCTCCGGAAGCCAGGAGGACCTTCTATGTGCCCCAGTCTTCTTGAAGGT
GGCTACCCAGCCTATGCAAGCGGGGGTGGCTACGTATTGCCGGGCGCCTGGCACCCCTGG
CTGCTGCGGGCGGCAGCCCGTGTGGCACCCCTTCCCCTTGAGGACGTCTACACTGGCCTT
TGCATCCGAGCCCTGGGCTGGTGGCCAGGCCACCCAGGCTTCCTCACAGCCTGGCCA
GCAGACCGCACTGCGGACCACTGTGCTTTCCGCAACCTGCTGCTGGTACGGCCCTGGGC
CCCCAGGCCAGCATTGGGCTCTGGAAACAAGTCAAGACCCAAGGCTCCAGTGTGACTC
TCATTGGGGAGGGCGGAGGTGCTGACCTGGCCCCGGCCCTGGCCTGGGCCTCTGGGGCG
GCCCTGGCTCAGCCCTCCTTCCAGGTCTTGATGGGAGGGAGGAGGGCCAGAAGCTGG
ACAACCTTAAGCCACTCCTTGGCCTCCCCAGCCAGGTGAGTGAGCTAT

A NOV81b polypeptide (SEQ ID NO:270) encoded by SEQ ID NO:269 is 397 amino acids in length and is presented using the one-letter amino acid code in Table 81D. The Psort profile for NOV81b predicts that this sequence has a signal peptide and is likely to be localized to lysosomes with a certainty of 0.8650, to the exterior of the cell with a certainty of 0.8190. The Signal P predicts a likely cleavage site for a NOV81b peptide is between positions 34 and 35, *i.e.*, at the dash in the sequence SAK-YP.

Table 81D. NOV81b Polypeptide Sequence (SEQ ID NO:270)

MRCPKCLLSALLTLGLKVYIEWTSESRLSKAYPSPRGTTPPSPTPANPEPTLPANLST
RLGQTIPLPFAYWNQQWRLGSLPSGDSTETGGCQAWGAAAATEIPDFASHPKDLRRFLL
SAACRSFPQWLPGGGSGVSSCDTDPYLLAVKSEPRFAERQAVRETWGSFAPGIRL
LFLLGSPVGEAGPDLDSLVAWESRRYSDDLWDFLDVFPNQTLDLALLLAWLGRHCPTVS
FVLRAQDDAFVHTPALLAHLRALPPASARSLYLGEVFTQAMPLRKPGGPFYVPESFFEGG

YPAYASGGGYVIAGRLAPWLLRAAARVAPFPFEDVYTGLCTRALGLVPPQAHGFLTAWPA
DRTADHCAFRNLLLVRPLGFPQASIRLWKQLQDPRLLQC

A BLAST analysis of NOV81 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV81 had high
5 homology to other proteins as shown in Table 81E.

Table 81E. BLASTX results from PatP database for NOV81					
Sequences producing High-scoring Segment Pairs:			High Score	Smallest Sum Probability P(N)	
patp:AAB03619	Human beta-1,3-galactosyltransferase Znssp2		2130	2.4e-220	
patp:AAB03620	Murine beta-1,3-galactosyltransferase Znssp		1585	1.4e-162	
patp:AAM41987	Human polypeptide		1528	1.5e-156	
patp:AAE05767	Human secreted protein (SECP)		641	1.5e-62	
patp:AAM40201	Human polypeptide		641	1.5e-62	

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 647 of 1088 bases (59%) identical to a gb:GENBANK-
10 ID:AP001754|acc:AP001754.1 mRNA from *Homo sapiens* (genomic DNA, chromosome 21q, section 98/105). The full amino acid sequence of the protein of the invention was found to have 127 of 343 amino acid residues (37%) identical to, and 194 of 343 amino acid residues (56%) similar to, the 397 amino acid residue ptnr:TREMBLNEW-ACC:AAD09763 protein from *Mus musculus* (Mouse) (BETA-1,3-N-ACETYLGLUCOSAMINYLTRANSFERASE
15 (EC 2.4.1.149)). NOV81 also has homology to the other proteins shown in the BLASTP data in Table 81F.

Table 81F. NOV81 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14290592 gb AAH09075.1 AAH09075 (BC009075)	beta-1,3-N-acetylglucosaminyltransferase 1 [<i>Mus musculus</i>]	397	126/350 (36)	192/350 (54)	4e-58
gi 16973463 gb AAL32299.1 AF321831_1 (AF321831)	beta-3-galactosyltransferase [<i>Danio rerio</i>]	406	128/346 (36)	186/346 (52)	1e-57

gi 9938024 ref NP_058584.2 (NM_016888)	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1; beta-1,3-N-acetylglucosaminyltransferase; beta-1,3-N-acetylglucosaminyltransferase 1; UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 6 [Mus musculus]	397	125/350 (35)	141/350 (53)	2e-57
gi 9845238 ref NP_006568.2 (NM_006577)	beta-1,3-N-acetylglucosaminyltransferase bGnT-1; beta3gal-T5 gene; beta-1,3-N-acetylglucosaminyltransferase bGnT-2 [Homo sapiens]	397	121/351 (34)	187/351 (52)	6e-56
gi 9664889 gb AAF97254.1 AF288209_1 (AF288209)	beta galactosyltransferase bGalt7 [Homo sapiens]	393	121/351 (34)	187/351 (52)	6e-56

This BLASTP data is displayed graphically in the ClustalW in Table 81G. A multiple sequence alignment is given, with the NOV81 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 81F.

Table 81G. ClustalW Alignment of NOV81	
NOV81a	(SEQ ID NO:268)
NOV81b	(SEQ ID NO:270)
gi 14290592	(SEQ ID NO:704)
gi 16973463	(SEQ ID NO:705)
gi 9938024	(SEQ ID NO:706)
gi 9845238	(SEQ ID NO:707)
gi 9664889	(SEQ ID NO:708)
	10 20 30 40 50
NOV81a	MRCPKCLLCISALLTLGLKVVLEWTSESRLSKAYPSPRGTPPSPTPANP
NOV81b	MRCPKCLLCISALLTLGLKVVLEWTSESRLSKAYPSPRGTPPSPTPANP
gi 14290592	MSVGRRRVKLLGILMMANVFIIYLIVEVSKNSSQDKNGKGGVILPKEKFWK
gi 16973463	MQGP RR--KVKVMAMMTMVFLFIVVEVSRNAGKSSSRNKSLVPLKRFWA
gi 9938024	MSVGRRRVKLLGILMMANVFIIYLIVEVSKNSSQDKNGKGGVILPKEKFWK
gi 9845238	MSVGRRRVKLLGILMMANVFIIYLIVEVSKSSSQEKNGKGEVILPKEKFWK
gi 9664889	-MVS R--S-LVGLMMANVFIIYLIVEVSKSSSQEKNGKGEVILPKEKFWK
	60 70 80 90 100
NOV81a	EPILPANLSTRLGQTIP-LPFAYWNQQQWRLGSLPSGDS-TETGGCQAWG
NOV81b	EPILPANLSTRLGQTIP-LPFAYWNQQQWRLGSLPSGDS-TETGGCQAWG
gi 14290592	PESTPRAYWNREOEKLNRYNPILNRYANOTGELATSPN-TSHLSYCEPD
gi 16973463	KDLPSDAYWNRRQQQINYNINRNLEKLYNTIDNLPDWLNDTVSLDSCDPD
gi 9938024	PESTPRAYWNREOEKLNRYNPILNRYANOTGELATSPN-TSHLSYCEPD
gi 9845238	ISTPPEAYWNREOEKLNRYNPILSMLTNOTGEAGRLSN-ISHLNYCEPD
gi 9664889	ISTPPEAYWNREOEKLNRYNPILSMLTNOTGEAGRLSN-ISHLNYCEPD
	110 120 130 140 150
NOV81a	AAAATEIPDFASYPKDLRRFLLSAACRSHPQWLPGGGGGQVSSCSDDTDVP

NOV81b	AAAATEIPDFASHPKDLRRFLLSAACRSFPQWLPGGGGSQVSSCSDDTDVP
gi 14290592	STVMTAVTDFFNNLPDRFKDFLLYLRCRNVSLLLID-----OPKKCAKK--P
gi 16973463	YRVTTQVKDYNLSLPDRFKDFLLYMCRSYPLVVD-----QPNICKKQ--P
gi 9938024	STVMTAVTDFFNNLPDRFKDFLLYLRCRNVSLLLID-----OPKKCAKK--P
gi 9845238	LRVTSVVTGFNNLPDRFKDFLLYLRCRNVSLLLID-----QPKCAKK--P
gi 9664889	LRVTSVVTGFNNLPDRFKDFLLYLRCRNVSLLLID-----QPKCAKK--P

	160	170	180	190	200
NOV81a	YLLLA	VKSE	PGRFA	ERQAV	RETWG
NOV81b	YLLLA	VKSE	PGRFA	ERQAV	RETWG
gi 14290592	FLLLA	IKSL	IPHFAR	ROAIRE	SWGRET
gi 16973463	FLLLA	IKSL	IPHFAR	ROAIRE	SWGRET
gi 9938024	FLLLA	IKSL	IPHFAR	ROAIRE	SWGRET
gi 9845238	FLLLA	IKSL	IPHFAR	ROAIRE	SWGRET
gi 9664889	FLLLA	IKSL	IPHFAR	ROAIRE	SWGRET

	210	220	230	240	250
NOV81a	DLDSL	VAWES	RRYS	DL	LLWDL
NOV81b	DLDSL	VAWES	RRYS	DL	LLWDL
gi 14290592	DLSD	MLKF	ESDK	HODIL	MWNYR
gi 16973463	DLSD	MLKF	ESDK	HODIL	MWNYR
gi 9938024	DLSD	MLKF	ESDK	HODIL	MWNYR
gi 9845238	DLSD	MLKF	ESDK	HODIL	MWNYR
gi 9664889	DLSD	MLKF	ESDK	HODIL	MWNYR

	260	270	280	290	300
NOV81a	RAQDD	AFVHT	PALLA	HLRAL	PPASAR
NOV81b	RAQDD	AFVHT	PALLA	HLRAL	PPASAR
gi 14290592	KGDD	DFVNT	THILN	YLNLS	LSKSKA
gi 16973463	KGDD	DFVNT	THILN	YLNLS	LSKSKA
gi 9938024	KGDD	DFVNT	THILN	YLNLS	LSKSKA
gi 9845238	KGDD	DFVNT	THILN	YLNLS	LSKSKA
gi 9664889	KGDD	DFVNT	THILN	YLNLS	LSKSKA

	310	320	330	340	350
NOV81a	ESFTE	CGYP	PAYAS	GCGY	VIAGRL
NOV81b	ESFTE	CGYP	PAYAS	GCGY	VIAGRL
gi 14290592	EVFY	TGVY	PPYAG	GCGFL	YSGPL
gi 16973463	ESME	VCMY	PAYAG	GCGFL	YSGPL
gi 9938024	EVFY	TGVY	PPYAG	GCGFL	YSGPL
gi 9845238	EVVY	SGLY	PPYAG	GCGFL	YSGPL
gi 9664889	EVVY	SGLY	PPYAG	GCGFL	YSGPL

	360	370	380	390	400
NOV81a	LGLV	POAHP	PGFLT	AWPADR	-TADH
NOV81b	LGLV	POAHP	PGFLT	AWPADR	-TADH
gi 14290592	LGLV	PEKH	KGFR	TFDIE	EKNK
gi 16973463	MGLA	PEKH	KGFR	TFDIE	EKNK
gi 9938024	LGLV	PEKH	KGFR	TFDIE	EKNK
gi 9845238	LGLV	PEKH	KGFR	TFDIE	EKNK
gi 9664889	LGLV	PEKH	KGFR	TFDIE	EKNK

	410
NOV81a	PRLQC-----
NOV81b	PRLQC-----
gi 14290592	PNLKC-----
gi 16973463	PGTV
gi 9938024	PNLKC-----
gi 9845238	AHLKC-----
gi 9664889	AHLKC-----

Table 81H lists the domain description from DOMAIN analysis results against NOV81. This indicates that the NOV81 sequence has properties similar to those of other proteins known to contain this domain.

Table 81H. Domain Analysis of NOV81				
gnl Pfam pfam01762, Galactosyl_T, Galactosyltransferase. This family includes the galactosyltransferases UDP-galactose:2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase and UDP-Gal:beta-GlcNAc beta 1,3-galactosyltransferase. Specific galactosyltransferases transfer galactose to GlcNAc terminal chains in the synthesis of the lacto-series oligosaccharides types 1 and 2. SEQ ID NO:874				
CD-Length = 195 residues, 98.5% aligned Score = 77.4 bits (189), Expect = 1e-15				
NOV81:	162	AERQAVRETWGSP--APGIRL--LFLLGSPVGEAGPDLDLSLVAWESRRYSDLLLWDFLDV	217	
Sbjct:	1	ARRNAIRKTWMNQNSRGGRIKSLFLVG--LAALDGKLLKLVMEEARLYGDIIVVDLED	58	
NOV81:	218	PFNQTLKDLLLAWLGRHCPTVSFVLRAQDDAFVHTPALLAHLRALPPASARSLYLGEVF	277	
Sbjct:	59	YLNLTCLKTLTILLYVVSCKPNALIGKIDDDVFVNPDLNLSLLEREYIDPSPLSFYGYII	118	
NOV81:	278	TQAMPLRKPGGPFYVPESFF-EGGYPAYASGGGYVIAGRLAPWLLRAAARVAPFPFEDVY	336	
Sbjct:	119	KNGEPVRTKSKWYVPPTAYPCSNYPYLSGPFYILSRDAAPLILKASKHRRFIKIEDVL	178	
NOV81:	337	-TGLCIRALGLVPQ	349	
Sbjct:	179	ITGILALDLGISRI	192	

5

There are 2 known types of carbohydrate chains in the lacto series of oligosaccharides: type 1 chains, which contain the Gal(beta-1-3)GlcNAc linkage, and type 2 chains, which contain the topoisomer Gal(beta-1-4)GlcNAc. The biosynthesis of both types of chains is catalyzed by specific galactosyltransferases (GalTs), which transfer galactose (Gal) to N-acetylglucosamine (GlcNAc)-terminating chains. Beta-4-GalT enzymes (e.g., GGTB2; are the galactosyltransferases responsible for type 2 chain biosynthesis, while beta-3-GalTs are the type 1 elongating enzymes.

Kolbinger et al. (1998) searched an expressed sequence tag (EST) database with the amino acid sequence of a human beta-3-GalT, which they called beta-3-GalT1, and identified human brain cDNAs encoding a novel beta-3-GalT, which they named beta-3-GalT2. The deduced 422-amino acid beta-3-GalT2 protein has a predicted type II transmembrane topology with 5 potential N-glycosylation sites, and a predicted molecular mass of 49,202 Da. Beta-3-GalT2 shares 46% amino acid identity with beta-3-GalT1, but has a 17-amino acid extension at the carboxy terminus and longer cytoplasmic and stem regions. Beta-3-GalT2 directed the

synthesis of type 1 chains in mammalian cells and transferred Gal to GlcNAc- and Gal-terminating acceptors in enzymatic assays.

Northern blot analysis demonstrated strong expression of a 3.5-kb beta-3-GalT2 transcript, and weaker expression of a 2.8-kb transcript, in heart and brain. Amado et al. (1998) stated that a human beta-3-galactosyltransferase gene, called beta-3-GalT1 by them, was isolated from a melanoma cell line using a transfection-cloning strategy. Beta-3-GalT1 is a predicted 326-amino acid protein. By carrying out a BLAST search of an EST database with the beta-3-GalT1 coding sequence, Amado et al. (1998) identified cDNAs encoding 3 other beta-3-galactosyltransferases, beta-3-GalT2, beta-3-GalT3, and beta-3-GalT4. The sequences of the 4 predicted proteins share 29 to 42% identity and have several conserved short sequence motifs. All 4 appear to be evolutionarily related, since their coding regions are contained in a single exon.

Using an insect cell expression system, Amado et al. (1998) showed that beta-3-GalT1 and beta-3-GalT2 are UDP-galactose:beta-N-acetyl-glucosamine beta-1,3 galactosyltransferases with similar kinetic properties. Northern blot analysis revealed that beta-3-GalT1 is expressed as a 6.5-kb mRNA exclusively in brain. Hennet et al. (1998) identified a mouse beta-3-GalT1 homolog, designated beta-3-GalTI, and found that the coding region was contained in a single exon.

NOV81 is predicted to be expressed in at least the following tissues: colon, blood, and lymphocyte. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV81 is provided in Example 2.

The NOV81 nucleic acids and proteins of are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmune disease, allergies, immunodeficiencies, transplantation, graft versus host disease (GVHD), lymphoedema, anemia, ataxia-telangiectasia, autoimmune disease, immunodeficiencies, Hirschsprung's disease, Crohn's Disease, appendicitis as well as other diseases, disorders and conditions. NOV81 nucleic acids encoding the galactosyltransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a beta-1,3-galactosyltransferase-like protein includes the nucleic acid whose sequence is provided in Table 81A or 81C, or a

fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 81A or 81C while still encoding a protein that maintains its beta-1,3-galactosyltransferase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
5 nucleic acids whose sequences are complementary to the sequence of Table 81A or 81C, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications
10 include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 41% of the bases
15 may be so changed.

The novel protein of the invention includes the beta-1,3-galactosyltransferase-like protein whose sequence is provided in Table 81B or 81D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 81B or 81D while still encoding a protein that maintains its beta-1,3-
20 galactosyltransferase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 63% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic
25 methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV82

30 The disclosed NOV82 (alternatively referred to herein as CG56983-01) includes the 348 nucleotide sequence (SEQ ID NO:271) shown in Table 82A. A NOV82 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 9-11 and ends with a stop codon at nucleotides 321-322. The disclosed NOV82 maps to human chromosome X.

Table 82A. NOV82 Nucleotide Sequence (SEQ ID NO:271)

CTATCCCTATGGTGTGCGGTGTGCAGGCCGTGGCCTGCTGTGGCCATAGCACTTCTGGCTCTGCTGGTCTG
CCTGGGGGCGCTGGTTCGACACCTGCCCCATCAAACCCGAGGCTCCTGGCGAAGACGAGTCCCTGGAGGAG
CTGAGCCACTATTATGCTTCCTGTGCCACTACCTCAACGTGGTCACCAGACAGTTAATTTTCAGAGAGAA
ACCTACCAGACACCATTGTGTCCAAGGAAGTATTTTTCACAAGCACAAAGGAAAGACCTGTGAGGACACA
GAAGGAAGGTTGCCATCTGCAAGCCAAGGAGAGAAGCCTCTGAAAAAACCAACCTGCTGGCACCTTG

A NOV82 polypeptide (SEQ ID NO:272) encoded by SEQ ID NO:271 is 104 amino acids in length and is presented using the one-letter amino acid code in Table 82B. The Psort profile for NOV82 predicts that this sequence has a signal peptide and is likely to be localized at the exterior of the cell with a certainty of 0.8200. In alternative embodiments, a NOV82 polypeptide is located to the endoplasmic reticulum (membrane) with a certainty of 0.1000. The Signal P predicts a likely cleavage site for a NOV82 peptide is between positions 28 and 29, i.e., at the dash in the sequence VDT-CP.

Table 82B. NOV82 Polypeptide Sequence (SEQ ID NO:272)

MVSVCRPWPAVAIALALLVCLGALVDTCP IKPEAPGEDESLEELSHYYASLCHYLNVT
RQLISERNLPDTIVSKEVFFSTKERPVRTQKEGCHLQAKERSL

A BLAST analysis of NOV82 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV82 had high homology to other proteins as shown in Table 82C.

Table 82C. BLASTX results from PatP database for NOV82

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAE09439 Human sbghPYYa protein - <i>Homo sapiens</i>	329	1.7e-29
patp:AAB08020 Amino acid sequence of a human peptide yY	301	1.6e-26
patp:AAG75364 Human colon cancer antigen protein	293	1.1e-25
patp:AAY14602 Amino acid sequence of the baboon PY	221	4.7e-18
patp:AAY43334 Neuropeptide Y - Synthetic, 97 aa.	188	1.5e-14

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 217 of 288 bases (75%) identical to a gb:GENBANK-ID:HUMPYYP3|acc:D13902.1 mRNA from *Homo sapiens* (Human mRNA for peptide YY). The full amino acid sequence of the protein of the invention was found to have 62 of 94 amino acid residues (65%) identical to, and 73 of 94 amino acid residues (77%) similar to, the 97 amino acid residue ptnr:SWISSNEW-ACC:P10082 protein from *Homo sapiens* (Human)

(PEPTIDE YY PRECURSOR (PYY)). NOV82 also has homology to the other proteins shown in the BLASTP data in Table 82D.

Table 82D. NOV82 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 1172796 sp P10082 PYY_HUMAN	PEPTIDE YY PRECURSOR (PYY) (PEPTIDE TYROSINE TYROSINE)	97	62/94 (65)	73/94 (76)	2e-23
gi 131753 sp P10631 PYY_RAT	PEPTIDE YY PRECURSOR (PYY) (PEPTIDE TYROSINE TYROSINE)	97	61/94 (64)	72/94 (75)	8e-23
gi 4758982 ref NP_004151.1 (NM_004160)	peptide YY [<i>Homo sapiens</i>]	90	60/91 (65)	71/91 (77)	4e-22
gi 422871 pir S34569	peptide YY precursor (clone L2) - human (fragment)	90	60/91 (65)	71/91 (77)	4e-22
gi 422870 pir S34568	peptide YY precursor (clone L1) - human (fragment)	90	59/91 (64)	70/91 (76)	2e-21

- 5 This BLASTP data is displayed graphically in the ClustalW in Table 82E. A multiple sequence alignment is given, with the NOV82 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 82D.

Table 82E. ClustalW Alignment of NOV82	
NOV82	(SEQ ID NO:272)
gi 1172796	(SEQ ID NO:709)
gi 131753	(SEQ ID NO:710)
gi 4758982	(SEQ ID NO:711)
gi 422871	(SEQ ID NO:712)
gi 422870	(SEQ ID NO:713)
<div> <div>1020304050</div> <div> NOV82MVSVCRPWPAVAIALALLVCLGALVDTCPIKPEAPGEDESLEELSHYYA gi 1172796 MVFVRRPWPALTTVLLALLVCLGALVDAYPIKPEAPGEDASPEELNRYYA gi 131753 MVAVRRPWPVVMVAMLLVLLACLGLALVDAYPAKPEAPGEDASPEELSRYYA gi 4758982 MVFVRRPWPALTTVLLALLVCLGALVDAYPIKPEAPGEDASPEELNRYYA gi 422871 MVFVRRPWPALTTVLLALLVCLGALVDAYPIKPEAPGEDASPEELNRYYA gi 422870 MVFVRRPWPALTTVLLALLVCLGALVDAYPIKPEAPGEDASPEELNRYYA </div> </div>	
<div> <div>60708090100</div> <div> NOV82SLCHYLNVTROLISERNLPDTIVSKEVFFSTKERPVRTQKECCHLOAK gi 1172796 SLRHYLNLVTRQRYGKRDGPDITLLSKTFFPDGEDRPVRSRSEGPDLW-- gi 131753 SLRHYLNLVTRQRYGKREVPAALESKLLFTDDSENLPFRSRPEGVDQW-- gi 4758982 SLRHYLNLVTRQRYGKRDGPDRLLSKTFFPDGEDRPVRSRSEGPDLW-- gi 422871 SLRHYLNLVTRQRYGKRDGPDITLLSKTFFPDGEDRPVRSR----- gi 422870 SLRHYLNLVTRQRYGKRDGPDRLLSKTFFPDGEDRPVRSR----- </div> </div>	
NOV82	ERSL

gi	1172796	----
gi	131753	----
gi	4758982	----
gi	422871	----
gi	422870	----

Table 82F lists the domain description from DOMAIN analysis results against NOV82. This indicates that the NOV82 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 82F. Domain Analysis of NOV82			
gnl Pfam pfam00159, hormone3, Pancreatic hormone peptide SEQ ID NO: 875			
CD-Length = 36 residues, 91.7% aligned			
Score = 43.1 bits (100), Expect = 8e-06			
NOV82:	30	PIKPEAPGEDESLEELSHYYASLCHYLNVTTRQ	62
		P KPE PG+D S E+L+ Y +L Y+N++TR	
Sbjct:	2	PSKPEYPGDDASPEDLAQYLRALRQYINLITRP	34

Pancreatic hormone (PP) is a peptide synthesized in pancreatic islets of Langherhans, which acts as a regulator of pancreatic and gastrointestinal functions. The hormone is produced as a larger propeptide, which is enzymatically cleaved to yield the mature active peptide, which is 36 amino acids in length and has an amidated C-terminus. The hormone has a globular structure with residues 2-8 forming a left-handed poly-proline-II-like helix, residues 9-13 a beta turn, and 14-32 an alpha-helix, held close to the first helix by hydrophobic interactions. Unlike glucagon, another peptide hormone, the structure of pancreatic peptide is preserved in aqueous solution. Both N- and C-termini are required for activity: receptor binding and activation functions may reside in the N- and C-termini respectively.

PYY is secreted from endocrine cells in the lower small intestine, colon, and pancreas. It acts on the gastrointestinal tract as an inhibitor of gastric acid secretion, gastric emptying, digestive enzyme secretion by the pancreas, and gut motility (Leiter et al., 1987). A related peptide, pancreatic polypeptide is secreted only by cells within the endocrine and exocrine pancreas and specifically inhibits the secretion of enzymes and bicarbonate from the exocrine pancreas. A third member of this gene family is neuropeptide Y.

Each of these proteins are synthesized with a signal peptide sequence followed by a 36-amino acid active peptide and a carboxyterminal peptide. During maturation, the signal and carboxyterminal peptides are cleaved and a common carboxyterminal tyrosine in the mature peptide is amidated. Hort et al. (1995) cloned the human PYY gene by screening a genomic library with a PCR product produced from the rat locus. The gene contains 4 exons spanning

about 1.2-kb of DNA. Exon 1 represents 5-prime untranslated sequence and is 75% identical to the comparable rat sequence.

PYY and PPY are about 10-kb apart and are mapped them by fluorescence in situ hybridization to 17q21. Based on a comparison of the 3 gene sequences, it has been concluded that NPY and PYY are the result of a gene duplication event, and that a subsequent tandem duplication produced the PPY gene. Pancreatic polypeptide, peptide tyrosine-tyrosine (PYY), and neuropeptide tyrosine (NPY), three members of a family of structurally related peptides, are mainly expressed in the endocrine pancreas, in endocrine cells of the gut, and in the brain, respectively. Synthetic human PYY prepared using a solid-phase synthetic technique was found to be structurally identical to the natural peptide.

NOV82 is predicted to be expressed in at least the following tissues: endothelium, esophageal carcinoma. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV82 is provided in Example 2.

The novel nucleic acids of the invention encoding a PEPTIDE YY -like proteins includes the nucleic acid whose sequence is provided in Table 82A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 82A while still encoding a protein that maintains its PEPTIDE YY -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 82A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 25% of the bases may be so changed.

The novel protein of the invention includes the PEPTIDE YY -like protein whose sequence is provided in Table 82B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 82B while

still encoding a protein that maintains its PEPTIDE YY -like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 35% of the amino acid residues may be so changed.

The NOV82 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, esophageal carcinoma, inflammatory bowel disease, diverticular disease, as well as other diseases, disorders and conditions. NOV82 nucleic acids encoding the peptide YY-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV83

The disclosed NOV83 (alternatively referred to herein as CG56890-01) includes the 1701 nucleotide sequence (SEQ ID NO:273) shown in Table 83A. A NOV83 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 19-21 and ends with a stop codon at nucleotides 1669-1671. The disclosed NOV83 maps to human chromosome 17.

Table 83A. NOV83 Nucleotide Sequence (SEQ ID NO:273)

```
TGCGCCCCGTGCTCAGCCATGGTGGACATGGGGGCCCTGGACAACCTGATCGCCAACACCGCCTACCTGC
AGGCCCGGAAGCCCTCGGACTGCGACAGCAAAGAGCTGCAGCGGCGGGCGGTAGCCTGGCCCTGCCCCGG
GCTGCAGGGCTGCGCGGAGCTCCGCCAGAAGCTGTCCCTGAACCTCCACAGCCTGTGTGAGCAGCAGCCCC
ATCGGTGCGCCGCTCTTCCGTGACTTCCCTAGCCACAGTGCCACGTTCCGCAAGGCGGCAACCTTCCTAG
AGGACGTGCAGAACTGGGAGCTGGCCGAGGAGGGACCCACCAAAGACAGCGCGCTGCAGGGGCTGGTGGC
CACTTGTGCGAGTGCCCCCTGCCCGGGGAACCCGCAACCCTTCTCAGCCAGGCCGTGGCCACCAAGTGC
CAAGCAGCCACCCTGAGGAAGAGCGAGTGGCTGCAGTGACGCTGGCCAAGGCTGAGGCCATGGCTTCT
TGCAAGAGCAGCCCTTTAAGGATTTCTGTGACCAGCGCCTTCTACGACAAGTTTCTGCAGTGGAACCTCTT
CGAGATGCAACCAGTGTGACACAAGTACTTCACTGAGTTTCAAGTGTCTGGGGAAGGTTGGTTTTGGGGAG
GTAAAAAACTGGGAAGATGTATGCCCTGTAAGAACTGGACAAGAAGCGGCTGAAGAAGAAAGGTGGCG
AGAAGATGGCTCTCTTGAAAAGGAAATCTTGAGAAGGTGAGCAGCCCTTTCATTGTCTCTCTGGCCTA
TGCCTTTGAGAGCAAGACCCATCTCTGCCTTGTCTATGAGCCTGATGAATGGGGGAGACCTCAAGTTCAC
ATCTACAACGTGGGCACGCGTGGCCTGGACATGAGCCGGGTGATCTTTTACTCGGCCAGATAGCCTGTG
GGATGCTGACCTCCATGAACTCGGCATCGTCTATCGGGACATGAAGCCTGAGAATGTCTTCTGGATGA
CCTCGGCAACTGCAGGTTATCTGACCTGGGGCTGGCGGTGGAGATGAAGGGTGGCAAGCCCATACCCAG
AGGCAGGCTGGAACCAATGGTTACATGGTCTCTGAGATCCTAATGGAAAAGGTAAGTTATCTCTATCCTG
TGGACTGGTTTTGCCATGGGATGCAGCATTATGAAATGGTTGCTGGACGAACACCATTCAAAGATTACAA
GGAAAAGGTGAGTAAAGAGGATCTGAAGCAAAGAAGTCTGCAAGACGAGGTCAAATTCAGCATGATAAC
TTCACAGAGGAAGCAAAGATATTTGCAGGCTCTTCTTGGCTAAGAAACCAGAGCAACGCTTAGGAAGCA
GGAGAGAAAAGTCTGATGATCCAGGAACATCATTCTTTAAACGATCAACTTTCCTCGCCTGGAAGC
TGGCTTAATTGAACCCCATTTGTGCCAGACCCTTCAGTGGTTTATGCCAAAGACATCGCTGAAATTGAT
GATTTCTGTGAGGTTGCGGGGGGTGGAATTTGATGACAAAGATAAGCAGTTCTTCAAAACTTTGCGACAG
GTGCTGTTCTATAGCATGGCAGGAAGAAATTATAGAAACGGGACTGTTTGAGGAAGTGAATGACCCCAA
CAGACTACGGGTGTGAGGAGGGTAATTCATCCAAGTCTGGCGTGTGTTTATTGTAATGTCTCTC
TTTACCAGACAGGCAGCAGGA
```


A NOV83 polypeptide (SEQ ID NO:274) encoded by SEQ ID NO:273 is 550 amino acids in length and is presented using the one-letter amino acid code in Table 83B. The Psort profile for NOV83 predicts that this sequence is likely to be localized to the nucleus with a certainty of 0.9685. In alternative embodiments, a NOV83 polypeptide is located to microbodies with a certainty of 0.1317.

Table 83B. NOV83 Polypeptide Sequence (SEQ ID NO:274)

```
MVDMGALDNLIAANTAYLQARKPSDCDSKELQRRRRSLALPGLQGCAELRQKLSLNFHSLC
EQQPIGRRLFRDFLATVPTFRKAATFLEDVQNWELAEEGPTKDSALQGLVATCASAPAG
NPQPFLSQAVATKCQAATTEEEERVAAVTLAKAEAMAFLEQPPKDFVTSAFYDKFLQWKL
FEMQPVSDKYFTEFRVLGKGGFGEVKNTGKMYACKKLDKKRLKKKGGEKMALLEKEILEK
VSSPFIVSLAYAFESKTHLCLVMSLMNGGDLKFHIYNVGTGRGLMSRVIFYSAQIACGML
HLHELGIYVRDMKPENVLDDLGNCRLSDLGLAVEMKGGKIPITQRQAGTNGYMAPEILME
KVSYSYPVDWFAMGCSIYEMVAGRTPFKDYKEKVSKEDEKQRTLQDEVKQFQHDNFTTEAK
DICRLFLAKKPEQRLGSRREKSDDPRKHHFFKTINFPRLEAGLIEPPFVDPDPVSVYAKDI
AEIDDFSEVRGVEFDDKDKQFFKNFATGAVPIAWQEEIIETGLFEELNDPNRPTGCEEGN
SSKSGVCLLL
```

A BLAST analysis of NOV83 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV83 had high homology to other proteins as shown in Table 83C.

Table 83C. BLASTX results from PatP database for NOV83

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAU03502 Human protein kinase #2 - <i>Homo sapiens</i>	2812	1.3e-292
patp:AAV57085 Human rhodopsin kinase amino acid sequence	1255	1.3e-127
patp:AAV24423 GRK4 polymorphism GRK4-alpha protein	1248	7.0e-127
patp:AAV24424 GRK4 polymorphism GRK4-beta protein	1190	1.7e-124
patp:AAV24425 GRK4 polymorphism GRK4-gamma protein	1215	2.2e-123

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1359 of 1700 bases (79%) identical to a gb:GENBANK-ID:AF063016|acc:AF063016.1 mRNA from *Spermophilus tridecemlineatus* (*Spermophilus tridecemlineatus* G protein-coupled receptor kinase GRK7 mRNA). The full amino acid sequence of the protein of the invention was found to have 463 of 549 amino acid residues (84%) identical to, and 502 of 549 amino acid residues (91%) similar to, the 548 amino acid residue ptnr:SPTREMBL-ACC:Q9Z2G7 protein from *Spermophilus tridecemlineatus*

(Thirteen-lined ground squirrel) (G PROTEIN-COUPLED RECEPTOR KINASE GRK7).
NOV83 also has homology to the other proteins shown in the BLASTP data in Table 83D.

Table 83D. NOV83 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17026318 gb AAL33880.1 AF282269_1 (AF282269)	G protein-coupled receptor kinase 7 [<i>Homo sapiens</i>]	553	548/555 (98)	548/555 (98)	0.0
gi 17933490 gb AAL06241.1 (AY049726)	retina G protein-coupled receptor kinase 7 [<i>Bos taurus</i>]	552	470/555 (84)	503/555 (89)	0.0
gi 17026320 gb AAL33881.1 AF282270_1 (AF282270)	G protein-coupled receptor kinase 7 [<i>Sus scrofa</i>]	553	465/555 (83)	504/555 (90)	0.0
gi 4001826 gb AAC95001.1 (AF063016)	G protein-coupled receptor kinase GRK7 [<i>Spermophilus tridecemlineatus</i>]	548	463/554 (83)	502/554 (90)	0.0
gi 17443851 ref XP_067593.1 (XM_067593)	similar to G protein-coupled receptor kinase GRK7 (<i>H. sapiens</i>) [<i>Homo sapiens</i>]	692	326/350 (93)	328/350 (93)	e-172

- 5 This BLASTP data is displayed graphically in the ClustalW in Table 83E. A multiple sequence alignment is given, with the NOV83 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 83D.

Table 83E. ClustalW Alignment of NOV83	
NOV83	SEQ ID NO:274)
gi 17026318	SEQ ID NO:714)
gi 17933490	SEQ ID NO:715)
gi 17026320	SEQ ID NO:716)
gi 4001826	SEQ ID NO:717)
gi 17443851	SEQ ID NO:718)
<div> <div>1020304050</div> <div> <div>NOV83</div> <div>gi 17026318 </div> <div>gi 17933490 </div> <div>gi 17026320 </div> <div>gi 4001826 </div> <div>gi 17443851 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>-----</div> <div>-----</div> <div>MPGEYKVTAPETSARDTHASVQVEAYQGREVHREISTLLKKAALPRKSS</div> </div> </div>	
<div> <div>60708090100</div> <div> <div>NOV83</div> <div>gi 17026318 </div> <div>gi 17933490 </div> <div>gi 17026320 </div> <div>gi 4001826 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>-----</div> <div>-----</div> </div> </div>	

gi	17443851	RETSFLRTAGSAEQSALHGCHKDTRNMRTSVYKGSVPCSDVGCDKFELQR
		110 120 130 140 150
	
NOV83		-----
gi	17026318	-----
gi	17933490	-----
gi	17026320	-----
gi	4001826	-----
gi	17443851	GHRGFVSFSKSKPEPEPEFRPSLQPPPLAAVSRAPDATRAAGPGRRESES
		160 170 180 190 200
	
NOV83		-----
gi	17026318	-----
gi	17933490	-----
gi	17026320	-----
gi	4001826	-----
gi	17443851	EFGNEELESIFSGLENNQRLQGLSLRYCGLGPQSGRLGSGVISQSAICEL
		210 220 230 240 250
	
NOV83		-----
gi	17026318	-----
gi	17933490	-----
gi	17026320	-----
gi	4001826	-----
gi	17443851	FLDGNYLECSGALALLRPIAGFAETQGEDQPAPGSPDTGNPPQRLQGHRT
		260 270 280 290 300
	
NOV83		-----
gi	17026318	-----MVDMGALDNLIAANTAY
gi	17933490	-----MVDMGALDNLIAANTAY
gi	17026320	-----MVDMGALDNLIAANTAY
gi	4001826	-----MDMGALDNLIAANTAY
gi	17443851	HCKSLGRCLTREGKAASSPPALLCFPWECAPCSAMVDMGALDNLIAANTAY
		310 320 330 340 350
	
NOV83		LQARKPSDCDSKELQRRRRSLALPGLOGCAELRQKLSLNFHSLCEQQPIG
gi	17026318	LQARKPSDCDSKELQRRRRSLALPGLOGCAELRQKLSLNFHSLCEQQPIG
gi	17933490	LQARKSSDADSKELQRRRRSLMLPGPOSCEQLRQALATDFHSLCEQQPIG
gi	17026320	LQARKTSDADSKELQRRRRSLMLPGPOSCEQLRQAMPADFSLCEQQPIG
gi	4001826	LQARKT-SDSDRELQRRRRSLALPGLOGCAELRQSLSPHFHSLCEQQPIG
gi	17443851	LQARKPSDCDSKELQRRRR---LRG-----VRQKLSLNFHSLCEQQPIG
		360 370 380 390 400
	
NOV83		RRLFRDFLATVPFERKAATFLEDVQNWELAEEGPTKDSALQGLVATCASA
gi	17026318	RRLFRDFLATVPFERKAATFLEDVQNWELAEEGPTKDSALQGLVATCASA
gi	17933490	RRLFRDFLATVPAYQEBARGFLEEVSWEELAEEGPAKGSALQGLVATCASA
gi	17026320	RRLFRDFLATVPAYQEBAMGFLEEVSWEELAEEGPAKGSTLQGLVATCASA
gi	4001826	RRLFRDFLATVPKYSQAVAFLEDVQNWELAEEGPAKTSTLQGLVATCARD
gi	17443851	RRLFRDFLATVPFERKAATFLEDVQNWELAEEGPTKDSALQGLVATCASA
		410 420 430 440 450
	
NOV83		PAPGNPQPFLSQAVATKCQAATTEEERVAAVTLAKAEAMAFLOEQPFFKDF
gi	17026318	PAPGNPQPFLSQAVATKCQAATTEEERVAAVTLAKAEAMAFLOEQPFFKDF
gi	17933490	PVPGRPHPFPSPALVTKCQAATTDDEERASLVELAKAEVMAFLQDQPPREF
gi	17026320	PNPGQPHSFLSPALVTKCQAATTDDEERASLVEQAKAEAMAFLODQPPREF
gi	4001826	P---GPQSFLSQDLATKCAASTDEERKTLVEQAKAETMSFLQEQPFFKDF
gi	17443851	PAPGNPQPFLSQAVATKCQAATTEEERVAAVTLRKAEAMAFLOEQPFFKDF
		460 470 480 490 500
	
NOV83		MTSAFYDKFLQWKLFEQMPVSDKYFTEFRVLGKGGFG-----EVKNTGKM
gi	17026318	MTSAFYDKFLQWKLFEQMPVSDKYFTEFRVLGKGGFG-----EVKNTGKM

gi	17933490		LASPFYDKFLQWKVFEMQPVSDKYFTEFRVLGKGGFGEVCAVQVNTGKM
gi	17026320		LVSPFYDKFLQWKVFEMQPVSDKYFTEFRVLGKGGFGEVCAVQVNTGKM
gi	4001826		LASPFYDRFLQWKLFEMQPVSDKYFTEFRVLGKGGFGEVCAVQVNTGKM
gi	17443851		VTSAPFYDKFLQWKLFEMQPVSDKYFTEFRVLGKGGFGEVCAVQVNTGKM

		510	520	530	540	550
NOV83	
gi	17026318		YACKKLDKKRLKKKGGEKMALLEKEILEKVSSPFIVSLAYAFESKTHLCL			
gi	17933490		YACKKLDKKRLKKKGGEKMALLEKEILEKVSSPFIVSLAYAFESKTHLCL			
gi	17026320		YACKKLDKKRLKKKGGEKMALLEKEILEKVSSPFIVSLAYAFESKTHLCL			
gi	4001826		YACKKLDKKRLKKKGGEKMALLEKEILEKVSSPFIVSLAYAFESKTHLCL			
gi	17443851		YACKKLDKKRLKKKGGEKMALLEKEILEKVSSPFIVSLAYAFESKTHLCL			

		560	570	580	590	600
NOV83	
gi	17026318		VMSLMNGGDLKFHIYNVGTGRGLDMSRVIFYSAQIACGMLHLHLGLGIVYRD			
gi	17933490		VMSLMNGGDLKFHIYNVGTGRGLDMSRVIFYSAQIACGMLHLHLGLGIVYRD			
gi	17026320		VMSLMNGGDLKFHIYSVGERGLDMNRVIFYSAQITCGVLHLHSLGLGIVYRD			
gi	4001826		VMSLMNGGDLKFHIYNVGTGRGLDMSRVIFYTAQMTCGVLHLHSLGLGIVYRD			
gi	17443851		VMSLMNGGDLKFHIYNVGTGRGLDMSRVIFYSAQIACGMLHLHLGLGIVYRD			

		610	620	630	640	650
NOV83	
gi	17026318		MKPENVLLDDLGNCRSLDLGLAVEMKGGKPITQRAGTNGYMAPEILMEK			
gi	17933490		MKPENVLLDDLGNCRSLDLGLAVEMKGGKPITQRAGTNGYMAPEILMEK			
gi	17026320		MKPENVLLDDLGNCRSLDLGLAVQIQDGKPITQRAGTNGYMAPEILMEK			
gi	4001826		MKPENVLLDDLGNCRSLDLGLAVEVQDDKPITQRAGTNGYMAPEILMEK			
gi	17443851		MKPENVLLDDLGNCRSLDLGLAVEMKGGKPITQRAGTNGYMAPEILMEK			

		660	670	680	690	700
NOV83	
gi	17026318		VSYSYPVDWFAMGCSIYEMVAGRTPPKDYKEKVSKEDEKQRTLODEVKFO			
gi	17933490		VSYSYPVDWFAMGCSIYEMVAGRTPPKDYKEKVSKEDEKQRTLODEVKFO			
gi	17026320		ASYSYPVDWFAMGCSIYEMVAGRTPPKDYKEKVSKEDEKQRTLODEVKFO			
gi	4001826		ASYSYPVDWFAMGCSIYEMVAGRTPPKDYKEKVSKEDEKQRTLODEVKFO			
gi	17443851		AMCAHVVD-----V-ELBAG-----VKER			

		710	720	730	740	750
NOV83	
gi	17026318		HDNFTTEAKDICRLFLAKKPEQRLGSR-EKSDDPRKHFFFTINFPRLA			
gi	17933490		HDNFTTEAKDICRLFLAKKPEQRLGSR-EKSDDPRKHFFFTINFPRLA			
gi	17026320		HSNFTTEAKDICRLFLAKKPEQRLGSR-EKSDDPRKHFFFTINFPRLA			
gi	4001826		HOSFTTEAKDICRLFLAKKPEQRLGSR-EKSDDPRKHFFFTINFPRLA			
gi	17443851		HENFTTEAKDICRLFLAKKPEQRLGSR-EKADDPRKHFFFTINFPRLA			

		760	770	780	790	800
NOV83	
gi	17026318		GLIEPPFVPDPSVYAKDIAEIDDFSEVRGVEFDDKDKQFFKQFATGAVP			
gi	17933490		GLIEPPFVPDPSVYAKDIAEIDDFSEVRGVEFDDKDKQFFKQFATGAVP			
gi	17026320		GLVEPPFVPDPSVYAKDIAEIDDFSEVRGVEFDDKDKQFFKQFATGAVP			
gi	4001826		GLVDPFVPDPSVYAKDIAEIDDFSEVRGVEFDDKDKQFFKQFATGAVP			
gi	17443851		GLVEPPFVPDPSVYAKDIAEIDDFSEVRGVEFDDKDKQFFKQFATGAVP			

		810	820	830
NOV83	
gi	17026318		IAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL	
gi	17933490		IAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL	
gi	17026320		IAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL	
gi	4001826		IAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL	
gi	17443851		IAWQEEIIETGLFEELNDPNRPTGCEEGNSSKSGVCLLL	

Table 83F lists the domain description from DOMAIN analysis results against NOV83. This indicates that the NOV83 sequence has properties similar to those of other proteins known to contain this domain.

Table 83F. Domain Analysis of NOV83					
gnl Smart smart00220, S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases.					
Serine or threonine-specific kinase subfamily. SEQ ID NO: 876					
CD-Length = 256 residues, 100.0% aligned					
Score = 208 bits (530), Expect = 6e-55					
NOV83:	191	FTEFRVLGKGGFGEV-----KNTGKMYACKLDDKKRLKKKGGEKMALEKEILEKVSSPF	245		
		+ VLGKG FG+V K TGK+ A K + K++LKKK E L E +IL+K+ P			
Sbjct:	1	YELLEVLGKGAFGKVYLARDKKTGKLVAIKVIKKEKLKKKKRE-RILREIKILKKLDHPN	59		
NOV83:	246	IVSLAYAFESKTHLCLVMSLMNGGDLKFHIYNVGTRGLDMSRVIFYSAQIACGMLHLHEL	305		
		IV L FE L LVM GGDL + G L FY+ QI + +LH			
Sbjct:	60	IVKLYDVFEDDDKLYLVMEYCEGGDLFDLLKKRG--RLSEDEARFYARQILSALEYLHSQ	117		
NOV83:	306	GIVYRDMKPENVLLDDLGNCRLSDGLAVEMKGGKPITQRQAGTNGYMAPEILMEKVSYS	365		
		GI++RD+KPEN+LLD G+ +L+D GLA ++ G + GT YMAPE+L+ K Y			
Sbjct:	118	GIIHRDLKPENILLDSGHVKLADFGLAQQLDSGGTLLTTFVGTPEYMAPEVLLGK-GYG	176		
NOV83:	366	YPVDWFAMGCSIYEMVAGRTPFKDYKEKVSKEDEKQRTLQDEVKF--QHDNFTEEAKDIC	423		
		VD +++G +YE++ G+ PF + L ++ + F + EAKD+			
Sbjct:	177	KAVDIWSLGVILYELLTGKPPFPDDQ---LLALFKKIGKPPPPPPPEWKISPEAKDLI	233		
NOV83:	424	RLFLAKKPEQRLGSRREKSDDPRKHFF	451		
		+ L K PE+RL +++ +H FF			
Sbjct:	234	KKLLVKDPEKRLT-----AEEALEHPFF	256		

Serine/threonine protein kinases are an extensive family of enzymes that catalyzes the phosphorylation of serine or threonine residues on its target protein. Protein kinases share a conserved catalytic core common to both serine/ threonine and tyrosine protein kinases. This domain contains residues, which are specific to the distinct types of protein kinases

The S6/H4 kinase purified from human placenta catalyzes phosphorylation of the S6 ribosomal protein, histone H4, and myelin basic protein. In vitro activation of the p60 S6/H4 kinase requires removal of an autoinhibitory domain by mild trypsin digestion and autophosphorylation of the catalytic domain (p40 S6/H4 kinase). The two autophosphorylation/autoactivation sites contain the sequences SSMVGTPY (site 1) and SVIDPVPAPVGDSHVDGAAK (site 2) (SEQ ID NO:1381). These sequences identify S6H4 kinase as the rac-activated PAK65 (Martin, G. A., Bollag, G., McCormick, F. and Abo, A. (1995) EMBO J. 14, 1971-1978). Site 1 phosphorylation is most rapid, but activation does not occur until site 2 is autophosphorylated. The site 1 phosphorylation occurs by an intramolecular mechanism whereas site 2 autophosphorylation occurs by an intermolecular mechanism. A model is proposed in which phosphorylation of sites 1 and 2 occurs sequentially. The model proposes that trypsin treatment of the inactive holoenzyme removes

an inhibitory rac-binding domain which blocks MgATP access to the catalytic site. The pseudosubstrate domain at site 1 is autophosphorylated and subsequent bimolecular autophosphorylation at site 2 fully opens the catalytic site. Phosphorylation by a regulatory protein kinase may occur at site 2 in vivo.

5 Rapid regulation of G-protein-coupled receptors (GPCRs) involves agonist- promoted receptor phosphorylation by GPCR kinases (GRKs) . This process is followed by arrestin binding and transient receptor internalisation. It has been shown that beta-adrenergic receptor kinase (beta ARK-1 or GRK2) follows a similar pattern of internalisation upon agonist activation of beta(2)-adrenergic receptors (beta(2)AR) and that beta ARK expression levels
10 modulate receptor sequestration.

 Such studies indicate a functional relationship between receptor phosphorylation and sequestration, showing that beta ARK not only translocates from the cytoplasm to the plasma membrane in response to receptor occupancy, but also shares endocytic mechanisms with the beta(2)AR . These results suggest a role for beta ARK in the sequestration process, or
15 involvement of receptor internalisation in the intracellular trafficking of the kinase.

 NOV83 is predicted to be expressed in at least the following tissues: retina, spleen. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV83 is provided in
20 Example 2.

 The NOV83 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft vesus host disease, Von Hippel-Lindau (VHL) syndrome, diabetes, tuberous sclerosis, as well as other
25 diseases, disorders and conditions. NOV83 nucleic acids encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

 The novel nucleic acid of the invention encoding a G protein-coupled receptor kinase GRK7-like protein includes the nucleic acid whose sequence is provided in Table 83A, or a
30 fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 83A while still encoding a protein that maintains its G protein-coupled receptor kinase GRK7-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes

nucleic acids whose sequences are complementary to the sequence of Table 83A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 21% of the bases may be so changed.

The novel protein of the invention includes the G protein-coupled receptor kinase GRK7-like protein whose sequence is provided in Table 83B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 83B while still encoding a protein that maintains its G Protein-Coupled Receptor Kinase GRK7-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 16% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV84

The disclosed NOV84 (alternatively referred to herein as CG56912-01) includes the 2355 nucleotide sequence (SEQ ID NO:275) shown in Table 84A. A NOV84 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 11-13 and ends with a stop codon at nucleotides 2336-2338. The disclosed NOV84 maps to human chromosome 17.

Table 84A. NOV84 Nucleotide Sequence (SEQ ID NO:275)

```
AGGCTGGAGGATGGGGCAGCAGGTCCTAGCGGGGAGGGCAGGGGAGCCTCTGGGCAGGTACGGCCTGAC
GCCCCGGGTCCCTCCCGCCCCGCCAGGCTGACGGAGGACGAGGACGTGCGCGCCATGCTGCGGGGCTCCC
GGCTCCGCAAGATCCGCTCGCGCACGTGGCACAAAGGAGCGGCTGTACCGGCTGCAGGAGGACGGCCTGAG
CGTGTGGTTCCAGCGGCGCATCCCGCGTGCGCCATCGCAGCACATCGTCTCCGCCCTGACCCGGCCCTC
CTCTCAGTCTTCGTGCAGCACATCGAGGCGGTCCGCGAGGGCCACCACTCCGAGGGCCTGCGGCGCTTCG
GGGGTGCTTCGCGCCAGCGCGCTGCCTCACCATCGCCTTCAAGGGCCGCGGCAAGAACCTGGACCTGGC
GGCGCCACGGCTGAGGAAGCGCAGCGCTGGGTGCGCGGTCTGACCAAGCTCCGCGCGCGCCTGGACGCC
ATGAGCCAGCGGAGCGGCTAGACCAATATTGGATCCACTCCTATCTGCACCGGGCTGACTCCAACCAGG
ACAGCAAGATGAGCTTCAAGGAGATCAAGAGCCTGCTGAGAATGGTCAACGTGGACATGAACGACATGTA
```

CGCTACCTCCTCTTCAAGCAGGAGTGTGACCACTCCAACAACGACCGTCTAGAGGGGGCTGAGATCGAG
GAGTTCCTGCGGCGGCTGCTGAAGCGGCGGAGCTGGAGGAGATCTTCCATCAGTACTCGGCGGAGGACC
GCGTGCTGAGTGCCCTGAGCTGCTGGAGTTCCTGGAGGACAGGGCGAGGGCGCCACACTGGCCCG
CGCCAGCAGCTCATTCAGACCTATGAGCTCAACGAGACACCCTCTCTGCCACCCCTATGACACTGGAT
GGCTTCATGATGTACCTGTTGTGCGCGGAGGGGGCTGCCTTGGACAACACCCACACGTGTGTGTCCAGG
ACATGAACCAGCCCCCTGGCCACTACTTCTCTCTCCACACACCTATCTGACTGACTCCAGAT
CGGGGGGGCCAGCAGCAGCAGGCTATGTTAGGGCCTTGGCCAGGGATGCCGCTGCGTGGAGCTGGAC
TGCTGGGAGGGGCGAGGAGGGGAGCCGTCATCTATCATGGCCATACCCCTACCTCCAAGATTCTCTTCC
GGACGCTGTCCAAGCCGTGCGCGACCATGCTTACGCTGTCCCTTACCCTGTCTATCTATCCCTGGA
CAACCACGACGGGCTGGAGCAGCAGGCTGCCATGGCCCGCCACCTCTGCACCATCTGGGGGACATGCTG
GTGACACAGGCGCTGGACTCCCCAAATCCCGAGGAGCTGCCATCCCGAGAGCAGCTGAAGGGCCGGTCC
TGGTGAAGGAAAGAAGCTGCGCGCTGCTCGGAGCGAGGATGGCCGGGCTCTGTCCGATCGGGAGGAGGA
GGAGGAGGATGACGAGGAGGAAGAAGAGGAGGTGGAGGCTGCAGCGCAGAGGCGGCTGGCCAAGCAGATC
TCCCGGAGCTGTGCGCCCTGGCTGTGTAAGTCCACGCCACCCGCTGCGGACCCCTGCACCTGCCCCCA
ACGCCCCACAACCCCTGCCAGGTGAGCTCCCTCAGCGAGCGCAAAGCCAAGAACTCATTCCGGAGGCAGG
GAACAGCTTTGTGAGGCACAATGCCCGCAGCTGACCCGCGTGTACCCGCTGGGGCTGCGGATGAATCA
GCCAACTACAGTCCCCAGGAGATGTGGAATCGGGCTGTGAGCTGGTGGCCTTGAACCTCCAGACGCCAG
GCTACGAGATGGACCTCAATGCCGGGCGCTTCTAGTCAATGGGCAGTGTGGCTACGCTCTAAAACCTGC
CTGCCTGCGGCAACCTGACTCGACCTTTGACCCGAGTACCCAGGACCTCCCAAGAACCTCTCAGCATC
CAGGTGCTGACTGCACAGCAGCTGCCCAAGCTGAATGCCGAGAAGCCACACTCCATTGTGGACCCCTGG
TGCGCATTGAGATCCATGGGCTGCCCGCAGACTGTGCCCGGCGAGGAGTACTGCTGCTCAACAATGG
CTTCAACCCCGCTGGGGGCGAGACCTGCAGTTCCAGCTGCGGGCTCCGGAGCTGGCACTGGTCCGTTT
GTGGTGAAGATTATGAGCCACCTCCCCAATGACTTTGTGGGCCAGTTTACTGCTCTTAGCAGCC
TAAAGCAAGGTACCGCCACATACACTGCTTTCCAAGGACGGGGCTCACTGTACACAGCCACGCTCTT
CATCCAAATCCGCATCCAGCGCTCTGAGGGGCCACCTCACTCGC

A NOV84 polypeptide (SEQ ID NO:276) encoded by SEQ ID NO:275 is 775 amino acids in length and is presented using the one-letter amino acid code in Table 84B. The Psort profile for NOV84 predicts that this sequence is likely to be localized to the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV84 polypeptide is located to microbodies with a certainty of 0.3000.

Table 84B. NOV84 Polypeptide Sequence (SEQ ID NO:276)
MGQQVLAGEGRGASGQVRPDAPGPPAPPGLTEDEDVRAMLRGSRIRKIRSRTWHKERLYR LQEDGLSVWFQRRIPRAPSQHIVLRPDPALLSVFVQHIEAVREGHQSEGLRRFSGGAFAPA RCLTIAFKGRRKNLDLAAPTAEAAQRWVRGLTKLRARLDAMSQRERLDQYWIHSYLHRAD SNQDSKMSFKEIKSLLRMVNDMNDMYAYLLFKQECDSNDRLEGAEIEEFRLRLLRKP ELEEIFHQYSGEDRVLSAPELLEFLQDQEEGATLARAQQLIQTYELNETPSPATPMTLD GFMMYLLSPEGAAALDNHTCTVCFQDMNQPLAHYFISSSHNTYLTDSQIGGPSSTEAYVRAF AQGRCVELDCWEGPGGEPVIYHGHTLTSKILFRDVQAVRDHAFTLSPPYVILSLDNHD GLEQQAAMARHLCTILGDMLVLTQALDSPNPEELPSPEQLKGRVVLVKGKKLPAARSEDGRA LSDREEEEEDEEEEEVEAAQRRRLAKQISPELSALAVYCHATRLRLHLPAPNAPQPCQ VSSLSEKAKKLIREAGNSFVRHNRQLTRVYPLGLRMNSANYSPOEMWNSGCQLVALNF QTPGYEMDLNAGRFLVNGQCGYVLKPACLRQPDSTFDPEYGPPTTLTSLQVLTAAQLPK LNAEKPHSIVDPLVRIEIHGVPADCARQETDYVLNNGFNPRWGQTLQFQLRAPELALVRF VVEDYDATSPNDFVGQFTLPLSSLKQGYRHIHLLSKDGASLSPATLFIQIRIORS

A BLAST analysis of NOV84 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV84 had high homology to other proteins as shown in Table 84C.

Table 84C. BLASTX results from PatP database for NOV84
Smallest

Sequences producing High-scoring Segment Pairs:	High Score	Sum Probability P(N)
patp:AAG63220 Amino acid sequence of a human protein	3735	0.0
patp:AAB47516 Human phospholipase C, 16835 - <i>Homo sapiens</i>	3734	0.0
patp:AAV81394 Rat phospholipase C-delta-1 - <i>Rattus</i> sp.	1882	4.6e-194
patp:AAE10440 Novel human phospholipase protein #7	1783	1.4e-183
patp:AAE11925 Human CG121 (or C592) lipase protein #1	1182	6.9e-120

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 954 of 1415 bases (67%) identical to a gb:GENBANK-ID:OCPLCMR|acc:Z49747.1 mRNA from *Oryctolagus cuniculus* (*O.cuniculus* mRNA for phospholipase C). The full amino acid sequence of the protein of the invention was found to have 381 of 745 amino acid residues (51%) identical to, and 524 of 745 amino acid residues (70%) similar to, the 745 amino acid residue ptnr:SPTREMBL-ACC:Q60450 protein from *Cricetulus griseus* (Chinese hamster) (PHOSPHOLIPASE C-DELTA1). NOV84 also has

10

Table 84D. NOV84 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14715017 gb AAH10668.1 AAH10668 (BC010668)	Similar to phospholipase C, delta [<i>Homo sapiens</i>]	613	604/615 (98)	606/615 (98)	0.0
gi 17481372 ref XP_053638.2 (XM_053638)	hypothetical protein XP_053638 [<i>Homo sapiens</i>]	581	472/500 (94)	475/500 (94)	0.0
gi 2137061 pir PC4183	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC 3.1.4.11) delta-1 - Chinese hamster (fragment)	745	378/758 (49)	522/758 (67)	0.0
gi 8393981 ref NP_058731.1 (NM_017035)	phospholipase C-delta1 [<i>Rattus norvegicus</i>]	756	368/757 (48)	501/757 (65)	0.0
gi 9790167 ref NP_062650.1 (NM_019676)	phospholipase C, delta; PLC-delta 1; phospholipase C delta-1 [<i>Mus musculus</i>]	756	374/757 (49)	503/757 (66)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 84E. A multiple sequence alignment is given, with the NOV84 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 84D.

15

Table 84E. ClustalW Alignment of NOV84

NOV84	(SEQ ID NO:276)
gi 14715017	(SEQ ID NO:719)
gi 17481372	(SEQ ID NO:720)
gi 2137061	(SEQ ID NO:721)
gi 8393981	(SEQ ID NO:722)
gi 9790167	(SEQ ID NO:723)

	10	20	30	40	50
NOV84	MGQQVVLAGEGRGASQVRPDAPGPPAPPGLTDEEDVRAMLRCGRRLRKIRS				
gi 14715017	-----MLRGRRLRKIRS				
gi 17481372	-----GLQDDODLQALLKGSOLLKVKKS				
gi 2137061	-----MDSGRDFTLHGLQDDPDLOALLKGSOLLKVKKS				
gi 8393981	-----MDSGRDFTLHGLQDDPDLOALLKGSOLLKVKKS				
gi 9790167	-----MDSGRDFTLHGLQDDPDLOALLKGSOLLKVKKS				

	60	70	80	90	100
NOV84	RTWHKERLYRLOEDGLSVWFQRR-LPRAPSQHIVLRPDPALLSVFVQHIE				
gi 14715017	RTWHKERLYRLOEDGLSVWFQRR-LPRAPSQHIF-----FVQHIE				
gi 17481372	SSWRRERFYKLOEDCKTIWOESRKVMRSPESOLFS-----IEDIQ				
gi 2137061	SSWRRERFYKLOEDCKTIWOESRKVMRSPESOLFS-----IEDIQ				
gi 8393981	SSWRRERFYKLOEDCKTIWOESRKVMRSPESOLFS-----IEDIQ				
gi 9790167	SSWRRERFYKLOEDCKTIWOESRKVMRSPESOLFS-----IEDIQ				

	110	120	130	140	150
NOV84	AVREGHQSEGLRRFGGAFAPARCLTIAFKGRKKNLDLAAPTAEEAQRWVR				
gi 14715017	AVREGHQSEGLRRFGGAFAPARCLTIAFKGRKKNLDLAAPTAEEAQRWVR				
gi 17481372	EVRMGHRTEGLEKFARDIPEDRCFSIVFKDQNTLDLIAPSADACHWVQ				
gi 2137061	EVRMGHRTEGLEKFARDIPEDRCFSIVFKDQNTLDLIAPSADACHWVQ				
gi 8393981	EVRMGHRTEGLEKFARDIPEDRCFSIVFKDQNTLDLIAPSADACHWVQ				
gi 9790167	EVRMGHRTEGLEKFARDIPEDRCFSIVFKDQNTLDLIAPSADACHWVQ				

	160	170	180	190	200
NOV84	GLTKLRARLDAMSORERLDQYWIHSYLHRADSNODSKMSFKEIKSLLRMV				
gi 14715017	-----MSORERLD-HWIHSYLHRADSNODSKMSFKEIKSLLRMV				
gi 17481372	GLTKLRARLDAMSORERLD-HWIHSYLHRADSNODSKMSFKEIKSLLRMV				
gi 2137061	GLRKLIHHSQSGMDOROKLQ-HWIHSCLRKADKNKDNKMNFKELKDFLKEI				
gi 8393981	GLRKLIHHSQSGMDOROKLQ-HWIHSCLRKADKNKDNKMNFKELKDFLKEI				
gi 9790167	GLRKLIHHSQSGMDOROKLQ-HWIHSCLRKADKNKDNKMNFKELKDFLKEI				

	210	220	230	240	250
NOV84	NVDMDMYAYLLEFKQECDHSNNDRLGAEIEEFLRLLKRPLEEIEFHQY				
gi 14715017	NVDMDMYAYLLEFK-ECDHSNNDRLGAEIEEFLRLLKRPLEEIEFHQY				
gi 17481372	NVDMDMYAYLLEFK-ECDHSNNDRLGAEIEEFLRLLKRPLEEIEFHQY				
gi 2137061	NIOVDDSYARKLER-ECDHSQTDSDLEDEIEETFYKMLTORAEIDRAFEEA				
gi 8393981	NIOVDDGYARKLER-ECDHSQTDSDLEDEIEETFYKMLTORAEIDRAFEEA				
gi 9790167	NIOVDDSYARKLER-ECDHSQTDSDLEDEIEETFYRMLTORAEIDRAFEEA				

	260	270	280	290	300
NOV84	SGEDRVLSAPELLEFLEDC-GEEGATLARAQQLIQTYELNETPSPATPMT				
gi 14715017	SGEDRVLSAPELLEFLEDC-GEEGATLARAQQLIQTYELNETAKQHEIMT				
gi 17481372	SGEDRVLSAPELLEFLEDC-GEEGATLARAQQLIQTYELNETAKQHEIMT				
gi 2137061	AGSAETLSVEKLVTFLOHQOREEEAAGPALALSLETERYEPSETAKAQRQMT				
gi 8393981	AGSAETLSVERLVTFLOHQOREEEAAGPALALSLETERYEPSETAKAQRQMT				
gi 9790167	AGSAETLSVEKLVTFLOHQOREEEAAGPALALSLETERYEPSETAKAQRQMT				

	310	320	330	340	350
NOV84	LDGFMYYLLSPECAALDNTHTCVFQDMNOPLAHYFTSSSHNTYLTDSQIG				
gi 14715017	LDGFMYYLLSPECAALDNTHTCVFQDMNOPLAHYFTSSSHNTYLTDSQIG				
gi 17481372	LDGFMYYLLSPECAALDNTHTCVFQDMNOPLAHYFTSSSHNTYLTDSQIG				
gi 2137061	KDGFLMYLLSADGSAFSLAHRVYQDMDOPLSHYLVSSSHNTYLLDQLT				
gi 8393981	KDGFLMYLLSADGSAFSLAHRVYQDMDOPLSHYLVSSSHNTYLLDQLT				

gi 9790167	KDGFLLMYLLSADGNFSLAHRVYQDMNOPLSHYLVSSSHNTYILEDQLT
	360 370 380 390 400
NOV84	GPSSTEAYVRAFAQGCRCVELDCWEGPGGEPVIYHGHTLTSKILERDVVQ
gi 14715017	GPSSTEAYVRAFAQGCRCVELDCWEGPGGEPVIYHGHTLTSKILERDVVQ
gi 17481372	GPSSTEAYVRAFAQGCRCVELDCWEGPGGEPVIYHGHTLTSKILERDVVQ
gi 2137061	GPSSTEAYTRALCKGCRCLELDCWDGPNQEPVIYHGHTLTSKILFYDVL
gi 8393981	GPSSTEAYTRALCKGCRCLELDCWDGPNQEPVIYHGHTLTSKILFYDVL
gi 9790167	GPSSTEAYTRALCKGCRCLELDCWDGPNQEPVIYHGHTLTSKILFYDVL
	410 420 430 440 450
NOV84	AVRDHAFITLSPYPVILSLDNHDCLEQQAAMARHLCTILGDMVLVTCALDSP
gi 14715017	AVRDHAFITLSPYPVILSLDNHDCLEQQAAMARHLCTILGDMVLVTCALDSP
gi 17481372	AVRDHAFITLSPYPVILSLDNHDCLEQQAAMARHLCTILGDMVLVTCALDSP
gi 2137061	ATRDYAFKASPPYPVILSLDNHDCLEQQAAMARHLKAILGPMILDDPLDG-
gi 8393981	ATRDYAFKASPPYPVILSLDNHDCLEQQAAMARHLKAILGPMILDDPLDG-
gi 9790167	ATRDYAFKASPPYPVILSLDNHDCLEQQAAMARHLKAILGPMILDDPLDG-
	460 470 480 490 500
NOV84	NPEELPSPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
gi 14715017	NPEELPSPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
gi 17481372	NPEELPSPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
gi 2137061	VTMSLPSPPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
gi 8393981	VTTSLPSPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
gi 9790167	VTTSLPSPEQLKGRVLYKGGKLP---AARSEDCGRALSDREEEEDDDEEEE
	510 520 530 540 550
NOV84	EEVEAAAQRRLAK---QISPELSALAVYCHATRLRLTHPAPNAPQ-CQV
gi 14715017	EEVEAAAQRRLAK---QISPELSALAVYCHATRLRLTHPAPNAPQ-CQV
gi 17481372	EEVEAAAQRRLAK---QISPELSALAVYCHATRLRLTHPAPNAPQ-CQV
gi 2137061	EAVRSQVQKSKEDKLVPELSDMVIYCKSVHFGGFSSESTSGQAFYEM
gi 8393981	EAVRSQVQKSKEDKLVPELSDMVIYCKSVHFGGFSSESTSGQAFYEM
gi 9790167	EAVRSQVQKSKEDKLVPELSDMVIYCKSVHFGGFSSESTSGQAFYEM
	560 570 580 590 600
NOV84	SSLSERKAKKRLIREAGNSFVRHNRQLTRVYPLCLRMNSANYSPQEMWNS
gi 14715017	SSLSERKAKKRLIREAGNSFVRHNRQLTRVYPLCLRMNSANYSPQEMWNS
gi 17481372	SSLSERKAKKRLIREAGNSFVRHNRQLTRVYPLCLRMNSANYSPQEMWNS
gi 2137061	APSASAKPR---NSFGROELCOACCPADPRVPAG--AADELS--
gi 8393981	ASFSENRLRLQLQESGNSFVRHNVSHLSRIYPACGRTDSSNYSPVEMWNG
gi 9790167	ASFSENRLRLQLQESGNSFVRHNVSHLSRIYPACGRTDSSNYSPVEMWNG
	610 620 630 640 650
NOV84	GCQIVALNFQTPG-YEMDINAGRFLVNGQCGYVLKPAFLRDPDSTFDPP-
gi 14715017	GCQIVALNFQTPG-YEMDINAGRFLVNGQCGYVLKPAFLRDPDSTFDPP-
gi 17481372	GCQIVALNFQTPG-YEMDINAGRFLVNGQCGYVLKPAFLRDPDSTFDPP-
gi 2137061	GCQIVALNFQTPG-PEMDVYLGCFQDNGGCGYVLKPAFLRDPDSTFDPP-
gi 8393981	GCQIVALNFQTPG-PEMDVYLGCFQDNGGCGYVLKPAFLRDPDSTFDPP-
gi 9790167	GCQIVALNFQTPG-PEMDVYLGCFQDNGGCGYVLKPAFLRDPDSTFDPP-
	660 670 680 690 700
NOV84	EYPGEPR--TTLSTQVLTAAQQLPKLNABKPHSIVDPLVRIEIHGVPAQCA
gi 14715017	EYPGEPR--TTLSTQVLTAAQQLPKLNABKPHSIVDPLVRIEIHGVPAQCA
gi 17481372	---GEQC--RAIPSQWA-----VWLR--PKTCL
gi 2137061	LTOGPWWAQKRLRVRIISGQQLPKVNKSK-NSIVDEKVIIEIHGVGQDVA
gi 8393981	LTOGPWWAQKRLRVRIISGQQLPKVNKSK-NSIVDEKVIIEIHGVGQDVA
gi 9790167	LTOGPWWAQKRLRVRIISGQQLPKVNKSK-NSIVDEKVIIEIHGVGQDVA
	710 720 730 740 750
NOV84	ROETDYVLNNGFNPRWGQTLQFQRAPELALVRFVVEDYDATSNDFFVGC
gi 14715017	ROETDYVLNNGFNPRWGQTLQFQRAPELALVRFVVEDYDATSNDFFVGC

gi 17481372	PAAT-----
gi 2137061	SROTAVITNNGFNPWWDTEFEFEVAVPDLALVRFVVEDYDASSKNDFIGQ
gi 8393981	SROTAVITNNGFNPRWDMFEFEFEVTVPDLALVRFVVEDYDSSSKNDFIGQ
gi 9790167	SROTAVITNNGFNPRWDTFEFEFEVAVPDLALVRFVVEDYDSSSKNDFIGQ
	760 770 780
NOV84	FTLPLSSLKQGYRHHLLSKDCASLSPATLFTQIRIORS
gi 14715017	FTLPLSSLKQGYRHHLLSKDCASLSPATLFTQIRIORS
gi 17481372	-----
gi 2137061	STIPWNSLKQGYRHHVLLSKNCDQHPSATLFFVKISIQD-
gi 8393981	STIPWNSLKQGYRHHVLLSKNCDQHPSATLFFVKISIQD-
gi 9790167	STIPWNSLKQGYRHHVLLSKNCDLHPSATLFFVKISIQD-

Table 84F lists the domain description from DOMAIN analysis results against NOV84. This indicates that the NOV84 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 84F. Domain Analysis of NOV84				
gnl Smart smart00148, PLCXc, Phospholipase C, catalytic domain (part); domain X; Phosphoinositide-specific phospholipases C. These enzymes contain 2 regions (X and Y) which together form a TIM barrel-like structure containing the active site residues. Phospholipase C enzymes (PI-PLC) act as signal transducers that generate two second messengers, inositol-1,4,5-trisphosphate and diacylglycerol. The bacterial enzyme appears to be a homologue of the mammalian PLCs. SEQ ID NO: 877				
CD-Length = 145 residues, 93.1% aligned				
Score = 186 bits (471), Expect = 6e-48				
NOV84:	323	QDMNQPLAHYFISSSHNTYLTDSQIGGPSSTEAYVRAFAQGCRCVELDCWEGPGGEPVIY	382	
		QDM++PL+HYFI+SSHNTYLT Q+ G SS E Y++A GCRCVELDCW+GP GEPVIY		
Sbjct:	1	QDMSKPLSHYFINSSHNTYLTGKQLWGESSVEGYIQALKHGCRCVELDCWDGPDGEPVIY	60	
NOV84:	383	HGHTLTSLKILFRDVVQAVRDHAFTLSPYPVILSLDNHDGLEQQAAMARHLCTILGDMMLVT	442	
		HGHT T I +V++A++ AF SPYPVILSL+NH +QQA MA+ I GD+L T		
Sbjct:	61	HGHTFTLPIKLSEVLEAIKKFAFVTSYPYPVILSLENHCSPDQQAQMAQMFKEIFGDLLEYT	120	
NOV84:	443	QALDSPNPEELPSPEQ	458	
		S + E LPSPEQ		
Sbjct:	121	PPTTS-SLEYLPSPEQ	135	

Phosphoinositide-specific phospholipase C acts as a signal transducer that generates two messengers, diacylglycerol and inositol 1,4,5-trisphosphate, by hydrolyzing inositol phospholipids. Molecules belonging to the PLC family are divided into subfamilies, PLC-beta, PLC-gamma, and PLC-delta, whose amino acid sequences are highly conserved in two distinct regions designated X and Y. PLC-delta-1 is distinguished from PLC-gamma by lack of the SH2 and SH3 domains that are essential for activation of PLC-gamma by tyrosine protein kinases, and from PLC-beta by lack of the C-terminal region of PLC-beta that is responsible for binding and activation by G proteins.

10

Cheng et al. (1995) cloned cDNA for human PLC-delta-1 and localized the gene to chromosome 3 by means of a human/rodent somatic cell panel (Lyu et al., 1996). In the course of a large-scale sequencing analysis of genomic DNA in the vicinity of the homozygous deletion on chromosome 3p found in a lung cancer cell line, Ishikawa et al. (1997) found that the gene encoding phospholipase C, delta-1 (PLCD1) is located just distal to the region removed by the deletion. They found that the gene consists of 15 exons and spans about 22 kb. By fluorescence in situ hybridization, they localized the PLCD1 gene to 3p22-p21.3. Shimohama et al. (1998) examined the entire sequences corresponding to protein-coding exons 2-15 of the hamster PLC-delta-1 gene in genomic DNA derived from the leukocytes of 13 unrelated patients with early-onset sporadic Alzheimer disease. In 1 of these patients whose clinical features and course did not differ from those of the other 12 cases, they found a change of codon CGC (arg) to CAC (his), located in the pleckstrin homology domain of the PLCD1 gene. They stated that this was the first mutation found in the human PLC genes.

Site-directed mutagenesis of the glutathione-S-transferase (GST/PLCD1) fusion protein changing arg105 to his resulted in a 4-fold decrease in the affinity of specific binding and a reduction in hydrolyzing activity to about 40% of that of the wildtype enzyme. This remarkable loss of function could be interpreted in terms of a conformational change in the pleckstrin homology domain. Shimohama et al. (1998) found that the arg105-to-his mutation was present in heterozygous state in the patient with AD. The mutation was not found in DNA extracted from leukocytes of 23 unrelated patients with familial AD, 23 unrelated patients with early-onset sporadic AD, 46 unrelated patients with late-onset sporadic AD, and 456 nondemented control subjects. Thus the change did not appear to be a common polymorphism.

However, determination of the possible pathologic role required transgenic studies of the mutant gene to determine the role of the enzyme and the mutation and a search for other mutations in the pleckstrin homology domain of PLC genes in human subjects with genetic disorders. In vitro single point mutagenesis, inositol phospholipid hydrolysis, and substrate protection experiments were used to identify catalytic residues of human phosphatidylinositol-specific phospholipase C delta 1 (PLC delta 1) isolated from a human aorta cDNA library. Invariant amino acid residues containing a functional side chain in the highly conserved X region were changed by in vitro mutagenesis. Most of the mutant enzymes were still able to hydrolyze inositol phospholipid with activity ranging from 10 to 100% of levels in the wild type enzyme. Exceptions were mutants with the conversion of Arg338 to Leu (R338L), Glu341 to Gly (E341G), or His356 to Leu (H356L), which made the enzyme severely defective in hydrolyzing inositol phospholipid. Phospholipid vesicle binding

experiments showed that these three cleavage-defective mutant forms of PLC delta 1 could specifically bind to phosphatidylinositol 4,5-bisphosphate (PIP2) with an affinity similar to that of wild type enzyme. Western blotting analysis of trypsin-treated enzyme-PIP2 complexes revealed that a 67-kDa major protein fragment survived trypsin digestion if the wild type enzyme, E341G, or H356L mutant PLC delta 1 was preincubated with 7.5 microM PIP2, whereas if it was preincubated with 80 microM PIP2, the size of major protein surviving was comparable to that of intact enzyme. However, mutant enzyme R338L was not protected from trypsin degradation by PIP2 binding. These observations suggest that PLC delta 1 can recognize PIP2 through a high affinity and a low affinity binding site and that residues Glu341 and His356 are not involved in either high affinity or low affinity PIP2 binding but rather are essential for the Ca(2+)-dependent cleavage activity of PLC.

NOV84 is predicted to be expressed in at least the following tissues: aorta, brain, colon, foreskin, heart, muscle, placenta, stomach, uterus, whole embryo, brain, colon, eye, head and neck, lung, muscle, ovary, pancreas, placenta, skin, stomach, uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV84 is provided in Example 2.

The NOV84 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation as well as other diseases, disorders and conditions. NOV84 nucleic acids encoding the phospholipase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a Phospholipase C delta 1 -like protein includes the nucleic acid whose sequence is provided in Table 84A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 84A while still encoding a protein that maintains its Phospholipase C delta 1 -like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 84A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 33% of the bases may be so changed.

The novel protein of the invention includes the Phospholipase C delta 1-like protein whose sequence is provided in Table 84B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 84 while still encoding a protein that maintains its Phospholipase C delta 1-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 49% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV85

The disclosed NOV85 (alternatively referred to herein as CG56955-01) includes the 4091 nucleotide sequence (SEQ ID NO:277) shown in Table 85A. A NOV85 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 13-15 and ends with a stop codon at nucleotides 4078-4080.

Table 85A. NOV85 Nucleotide Sequence (SEQ ID NO:277)

ACTACTTGTGGAATGTCACCTCGGGGTATTTCAACAAGACAGGTACCTCTTGTGAAAGTCCGAAGTA ATTCTCTGAAAGCTCCTTCCACGCATGTCACAAAACCATCATTTAGCCAGAAATCATTGTCTTCTATGAG AGACCAAAGACCAGTAAATCACTTGCATCAGAACAGTCTGTTGAATCAGCAGACATGGGTAAAGGACTGAC AGTGCCCCCGATCAGCAAGTGGAGACTGGGAAATCCCCCTCTTTATCTGGAGCCTCTGCCAAGCCTGCCC CTCAGTCGAGTGAACCGTGGTACTTCAGATTAGAACTACCTGTCAGTCAAAGGAATCAAGATTAAAG TTTACAAGAGGCTGAACTGAGCAATCAGATACTTTAGATAATAAAGAAGCTGTCATCCTAAGGGAAAAA CCTCCATCTGGACGCCAGACACCGCAGCCTTTAAGGCATCAGTCTTACATCTTGGCAGTAAATGACCAGG AGACCGGGTCAGACACTACCTGCTGGCTGCCCAATGATGCACGTCGAGAGGTCCACATAAAAAGAATGGA GGAAGAAAAGCCTCGAGTACCAGTCCGCTGGCGATTCTTGGCTTCCATCCCATTATAGATGAACCA ACTAGCCCTAGCATTGATCATGATATTGCACATATCCCTGCCTCTGCTGTTATATCAGCCTCTACCTCTC AGGTCCCTCCATAGCAACAGTTCCCTCCTTGCCCTCACAACCTCAGCTCCATTAATTCGCCGTCAGCTCTC ACATGACCACGAATCTGTTGGCCCTCCTAGCCTGGATGCTCAGCCCAACTCAAAGACAGAAAGATCAAAA TCATATGATGAGGGTCTGGATGATTACAGAGAAGATGCAAATTGTCCTTTAAGCACGTATCTAGTCTGA AGGGAATCAAGATCGCAGACAGCCAAAAGTCATCAGAAGACTCTGGGTCCAGAAAAGATTCTTCCTCAGA GGTCTTCAGTGATGCTGCCAAGGAAGGGTGGCTTCATTTCCGACCCCTTGTCAACCGATAAGGGCAAGCGA

GTTGGTGAAGTATTCGGCCATGGAACAGATGTATGTTGTCCTTCGGGGTCATCACTTTACCTGTACA
 AAGATAAAAGAGAGCAGACCACTCCGTCTGAGGAAGAGCAGCCCATCAGTGTAAATGCTTGGTTGATAGA
 CATCTCTTACAGTGAGACCAAGAGGAAAAATGTGTTTCGACTCACCACGTCCGACTGTGAATGCCTGTTT
 CAGGCTGAAGACAGAGATGATATGCTAGCTTGGATCAAGACGATCCAGGAGAGCAGCAACCTAAACGAAG
 AGGACACTGGAGTCACTAACAGGGATCTAATTAGTCTGAAGAATAAAAGAATACAACAACTCTGATGAGCAA
 AGCAGAACAGTTGCCAAAAACACCTCGCCAGAGTCTCAGCATCAGGCAAACTTTGCTTGGTGCTAAATCA
 GAGCCAAAGACTCAAAGCCACACTCTCCGAAGGAAGAGTCCGAAAGGAACTTCTCAGTAAAGATGATA
 CCAGTCCCCCAAAGACAAAGGCACATGGAGAAAAGGCATTCCAAGTATCATGAGAAAGACATTTGAGAA
 AAAGCCAACTGTACAGGAACCTTCGGCGTCCGACTAGATGACTGCCACCAGCTCATACTAATCGGTAT
 ATTCATTAAATAGTTGACATATGTTGCAAATTAGTTGAAGAAAGAGTCTTGAATATACAGGTATTTATA
 GAGTTCCTGGAAATAATGCAGCCATCTCAAGTATGCAAGAAGAACTCAACAAGGGAATGGCTGATATTGA
 TATACAAGATGATAAATGGCGAGATTTGAATGTGATAAGCAGTTTACTAAAATCCTTCTTCAGAAAACCT
 CCTGAGCCTCTCTTCACAAATGATAAATATGCTGATTTTATTGAAGCCAATCGTAAAGAAGATCCTCTAG
 ATCGTCTGAAAACATTAAAAAGACTAATTCACGATTTGCCTGAACATCAATTATGAAACACTTAAAGTTCCT
 TTCAGCTCATCTGAAGACAGTGGCAGAAAATTTCAGAGAAAAAATAAGATGGAACCAAGAAACCTAGCA
 ATAGTGTGTTGGTCCCACCTTGTTCGAACATCAGAAGACAACATGACCCACATGGTCACCCACATGCCTG
 ACCAGTACAAGATTGTAGAAACGCTCATCCAGCACCATGACTGGTTTTTCACAGAAGAAGTGTGTAAGA
 GCCTCTTACAACAGTGCAGGAGGAAAGCACAGTAGACTCCAGCCAGTGCCAAACATAGATCATTTACTC
 ACCAACATTGGAAGGACAGGAGTCTCCCGAGGAGATGTATCAGATTGAGTACTAGTGAAGTCAACAAAAT
 CTAAGGGTCTTGGGGATCTGGAAGGATCAGTATAGCAGGGAAGTGTGTGTCTCCATCTTTGCAGC
 TGCTAGTCGCAAGAGGAAGAAGCCGAAAGAAAAGCACAGCCTAGCAGCTCAGAAGATGAAGTGGACAAT
 GTATTTTAAAGAAAGAAAATGTGGAACAGTGTCACAATGATACTAAAGAGGAGTCCAAAAAGAAAGTG
 AGACACTGGGCAGAAAACAGAAGATCATCATTGCCAAAGAAAACAGCACTAGGAAAGACCCACAGCAGAC
 AAAAGATGAAAAGATATCACTAGGAAAAGAGAGCAGCCTTCTGAAGAACCCTCACCACCACACAACCTCA
 AACACACAAGTACCAACTCTCAGCTGTGCTTTTGGCCATCTGAAAGAGAGCCCAAGTCACTTCTGG
 CACAGAAGTCTCTCCACCTTGAAGAGACAGGCTCTGACTCTGGCACTTTGCTCAGCAGCTCTTCCAGGC
 CTCCTTGCAAGGTTTTCCATGAAGAAATCAACCAGTCCAGAAACGAAACATAGCGAGTTTGTGGCCAAC
 GTCAGCACCATCACCTCAGATTATTCCACCACATCGTCTGCTACATACTTGACTAGCCTGGACTCCAGTC
 GACTGAGCCCTGAGGTGCAATCCGTGGCAGAGAGCAAGGGGACGAGGCAGATGACGAGAGAAGCGAACT
 CATCAGTGAAGGGCGGCTGTGGAACCGACAGCGAGAGCGAGTTTCCCGTGTTCCCACAGCCTTGACT
 TCAGAGAGGCTTTTCCGAGGAAAAGTGCAGAAAGTGAAGAGCAGCCGAGAGAAATCTGAAGGAAGTG
 AATTAAGTTGCACCGAGGGAAGTTTAAACATCAAGTTTAGATAGCCGAGACAGCTCTTCAGTTCCATAA
 ACTCATCGAATGTGATACTCTTCCAGGAAAAAATCAGCTAGATTCAAGTCAGATAGTGAAGTCTAGGA
 GATGCCAAGAAATGAGAAAGAAGCACCTTCGTTAACTAAAGTGTGTGATGTTATGAAAAAGGAAAGTCAA
 CTGGGAGTTTACTGACACCCACCAGAGGCGAATCCGAAAAACAGGAACCCACATGGAAGAACGAAATAGC
 AGATCGGTTAAAACTGAGACCCAGAGCCCTGCGGATGACATGTTTGGAGTAGGGAATCACAAAGTGAAT
 GCCGAGACTGCTAAAAGGAAAAGCATCCGGCGCAGACATACACTAGGAGGGCACAGAGATGCTACCGAAA
 TCAGCGTTTGAATTTTGGAAAGTGCATGAGCAGAGCGGGGAGAGAGAATCTGAACCTTCAGCTGTAAA
 CCGGTTAAACCAAAATGCTCAGCCAGGACCTTTCCATCTCAGACTGGCTGGCCAGGGAACGCTACGC
 ACCAGTACCTCTGACCTTAGCAGAGGAGAAATCGGAGATCCCCAGACAGAGAACCAGCACAGAGAAA
 TAGCCACGACCGACACACCTTTGTCTCTTCATTGCAACACAGGCAGTTCTTCCAGCACCTTGGCTTCAAC
 AAACAGGCCCTTCTTCCATACCACACAGTCACTGACCAATAAACGGAGAAAGCTTCCAGAACGTG
 AGCAAAATGCTAGTTCTGCAGCGAATGCCCAACCTCATAACTGTCTGAAACCCAGGCACTAAAGCAG
 AGTTTCATCCTGTCTTTAAACTGGGGGTAT

A NOV85 polypeptide (SEQ ID NO:278) encoded by SEQ ID NO:277 is 1355 amino acids in length and is presented using the one-letter amino acid code in Table 85B. The Psort profile for NOV85 predicts that this sequence is likely to be localized to the nucleus with a certainty of 0.7000. In alternative embodiments, a NOV85 polypeptide is located to the mitochondrial matrix space with a certainty of 0.3600, or to microbodies with a certainty of 0.3000.

Table 85B. NOV85 Polypeptide Sequence (SEQ ID NO:278)

MSLPRGISQDRSPLVKVRNSLSKAPSTHVTKPSFSQKSFVSMRDQRPVNLHQNLSLLNQ
 TWVRTDSAPDQVETGKSPSLSGASAKPAPQSSNAGTSDLELPVSQRNQDLSLQEAETE
 QSDTLDNKEAVILREKPPSGRQTPQLRHQSYILAVNDQETGSDTTCLPNDARREVHIK
 RMEERKASSTSPPGDSLASIPFIDEPTSPSIDHIAHIPASAVISASTSQVPSIATVPFC
 LTTSAPLIRRLSHDHESVGPSPSLDAQPNKTERSKSYDEGLDDYREDAKLSFKHVSSLK
 GIKIADSQKSSDSGRKDSSEVFSDAAKEGWLHFRPLVTDKGRVGSIRPWKQMYVV

LRGHSLYLYKDKREQTTTPSEEEQPISVNACLIDISYSETKRKNVFRLLTSDCECLFQAED
RDDMLAWIKTIQESSNLNEEDTGVNDRDLISRIKEYNNLMSKAEQLPKTPRQSLSTRQT
LLGAKSEPKTQSPHSPKEESERKLLSKDDTSPPKDKGTWRKGIPSIMRKTFEKKPTATGT
FGVRLDDCPPAHTNRYIPLIVDICCKLVEERGLEYTGIIYRVPGNNAAISSMQEELNKGMA
DIDIQDDKWRDLNVISSLLKSFFRKLPEPLFTNDKYADFIENRKEDPLDRLKTLKRLIH
DLPEHHYETLKFSAHLKTVANSEKKNKMEPRNLAIVFGPTLVRTSEDNMTHMVTMPD
QYKIVETLIQHHDWFFTEEGAEPLTTVQEEESTVDSQPVNIDHLLTNIGRTGVSPGDVS
DSATSDSTKSKGSWGSGKDQYSRELLVSSIFAAASRKRKKPKKEKAQPSSSEDELNDNVFFK
KENVEQCHNDTKEESKKESETLGRKQKIIIAKENSTRKDPSTTKDEKISLGKESTPSEEP
SPPHNSKHNSPTLSCRFAILKESPRSLAQKSSHLEETGSDSGTLLSTSSQASLARFSM
KKSTSPETKHSEFLANVSTITSYDSTSSATYLTSLDSSRLSPEVQSVAESKGDEADDER
SELISEGRPVETDSESEFPVFPTALTSERLFRGKLQEVTKSSRRNSEGSELSCTEGSLTS
SLDSRRQLFSSHKLIECDTLRSRKSARFKSDSGSLGDAKNEKEAPSLTKVFDVMKKGKST
GSLLTPTRGESEKQPTWTKIADRLKLRPRAPADDMFGVGNHKVNAETAKRKSIRRRHT
LGGHRDATEISVLNFWKVHEQSGERESELSAVNRLPKCSAQDLSISDWLARERLRTSTS
DLSRGEIGDPQTENPSTREIATDTPLSLHCNTGSSSSTLASTNRPLLSIPPQSPDQING
ESFQNVSKNASSAANAQPHKLSETPGTKAEFHPCL

A BLAST analysis of NOV85 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV85 had high homology to other proteins as shown in Table 85C.

5

Table 85C. BLASTX results from PatP database for NOV85		
Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P (N)
patp:AAB97911 Human G-protein activating protein	6012	0.0
patp:AAB41660 Human ORFX ORF1424 polypeptide sequence	4368	0.0
patp:AAM93705 Human polypeptide	852	3.9e-84
patp:AAU17101 Novel signal transduction pathway protein	618	6.3e-59
patp:AAB64387 Amino acid sequence of human intracellular	426	2.1e-38

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 3871 of 3875 bases (99%) identical to a gb:GENBANK-ID:AB037845|acc:AB037845.1 mRNA from *Homo sapiens* (mRNA for KIAA1424 protein).

10 The full amino acid sequence of the protein of the invention was found to have 1285 of 1287 amino acid residues (99%) identical to, and 1286 of 1287 amino acid residues (99%) similar to, the 1286 amino acid residue ptnr:SPTREMBL-ACC:Q9P2C3 protein from *Homo sapiens* (Human) (KIAA1424 PROTEIN). NOV85 also has homology to the other proteins shown in the BLASTP data in Table 85D.

15

Table 85D. NOV85 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7243229 dbj BAA92662.1 (AB037845)	KIAA1424 protein [<i>Homo sapiens</i>]	1286	1285/1287 (99)	1286/1287 (99)	0.0

gi 14736381 ref XP_038564.1 (XM_038564)	hypothetical protein XP_038564 [Homo sapiens]	1173	1172/1174 (99)	1173/1174 (99)	0.0
gi 7959263 dbj BAA96025.1 (AB040934)	KIAA1501 protein [Homo sapiens]	735	331/685 (48)	433/685 (62)	e-148
gi 14748447 ref XP_041405.1 (XM_041405)	hypothetical protein XP_041405 [Homo sapiens]	304	277/284 (97)	279/284 (97)	e-135
gi 12856667 dbj BAB30742.1 (AK017433)	KIAA1424 PROTEIN (FRAGMENT)~putative [Mus musculus]	385	301/384 (78)	328/384 (83)	e-127

This BLASTP data is displayed graphically in the ClustalW in Table 85E. A multiple sequence alignment is given, with the NOV85 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 85D.

Table 85E. ClustalW Alignment of NOV85	
NOV85	(SEQ ID NO:278)
gi 7243229	(SEQ ID NO:724)
gi 14736381	(SEQ ID NO:725)
gi 7959263	(SEQ ID NO:726)
gi 14748447	(SEQ ID NO:727)
gi 12856667	(SEQ ID NO:728)
<div> <div>1020304050</div> <div> <div>NOV85</div> <div> <div>gi 7243229 </div> <div>gi 14736381 </div> <div>gi 7959263 </div> <div>gi 14748447 </div> <div>gi 12856667 </div> </div> </div> </div>	
<div> <div>60708090100</div> <div> <div>NOV85</div> <div> <div>gi 7243229 </div> <div>gi 14736381 </div> <div>gi 7959263 </div> <div>gi 14748447 </div> <div>gi 12856667 </div> </div> </div> </div>	
<div> <div>110120130140150</div> <div> <div>NOV85</div> <div> <div>gi 7243229 </div> <div>gi 14736381 </div> <div>gi 7959263 </div> <div>gi 14748447 </div> <div>gi 12856667 </div> </div> </div> </div>	
<div> <div>160170180190200</div> <div> <div>NOV85</div> <div> <div>gi 7243229 </div> <div>gi 14736381 </div> <div>gi 7959263 </div> <div>gi 14748447 </div> </div> </div> </div>	

gi	12856667	210	220	230	240	250
NOV85		-----				MSLPRGISQDRSP		-----
gi	7243229	NSKTERSKSYDEGLDDYREDAKLSFKHVSSLKGIKIADSQKSSSEDSGRK						-----
gi	14736381	-----						-----
gi	7959263	-----						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		260	270	280	290	300
NOV85		-----						-----
gi	7243229	LVKVRSNSLKPSTHVTKPSFSQKSFVSMRDQRPVNLHQNSSLNQQTWV						-----
gi	14736381	DSSSEVFSDAAKEGWLHFRPLVTDKGRVGGSIIRPWKQMYVVLRGHSLYL						-----
gi	7959263	-----						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		310	320	330	340	350
NOV85		-----						-----
gi	7243229	RTDSAPDQQVETGKSPSLSGASAKPAPQSSSENAGTSDLELPVSQRNQDLS						-----
gi	14736381	YKDKREQTTPSEEEQPISVNACLIDISYSETKRKNVFRLTSDCECLFQA						-----
gi	7959263	-----						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		360	370	380	390	400
NOV85		-----						-----
gi	7243229	LQEAETEQSDTLDNKEAVILREKPPSGRQTPQPLRHQSYILAVNDQETGS						-----
gi	14736381	EDRDDMLAWIKTIQESSNLNEEDTGVTNRDLSRRIKEYNNLMSKAEQLP						-----
gi	7959263	-----						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		410	420	430	440	450
NOV85		-----						-----
gi	7243229	DTTCWLPNDARREVHIKRMEERKASSTSPPGDSLASIPFIDEPTSPSIDH						-----
gi	14736381	KTPRQSLSIHQTLGAKSEPKTQSPHSPKEESERKLLSKDDTSPPKDKGT						-----
gi	7959263	-----MEERKASSTSPPGDSLASIPFIDEPTSPSIDH						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		460	470	480	490	500
NOV85		-----						-----
gi	7243229	DIAHIPASAVISASTSQVPSIATVPPCLTTSAPLIRRQLSHDHESVGPPS						-----
gi	14736381	WRKGIPNSIMRKTFEKKPTATGTGFGVRLDDCPPAHNTNRYIPLIVDICCKLV						-----
gi	7959263	DIAHIPASAVISASTSQVPSIATVPPCLTTSAPLIRRQLSHDHESVGPPS						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		510	520	530	540	550
NOV85		-----						-----
gi	7243229	LDAQPNKTERSKSYDEGLDDYREDAKLSFKHVSSLKGIKIADSQKSSSE						-----
gi	14736381	EERGLEYTGIYRVPGNNAIISMQEELNKGMAIDIDIQDDKWRDLNVISSL						-----
gi	7959263	LDAQPNKTERSKSYDEGLDDYREDAKLSFKHVSSLKGIKIADSQKSSSE						-----
gi	14748447	-----						-----
gi	12856667	-----						-----
		560	570	580	590	600
NOV85		-----						-----
gi	7243229	SGSRKDSSEVFSDAAKEGWLHFRPLVTDKGRVGGSIIRPWKQMYVVLRG						-----
gi	12856667	LKSFRKLPPEPLFTNDKYADFIEANRKEDPLDRLKTLKRLIHDLPPEHHE						-----

gi	14736381	SGSRKDSSEVFSDAAKEGWLHFRPLVTDKGKRVGGSIRPWKQMYVVLRG
gi	7959263	-----
gi	14748447	-----
gi	12856667	-----
		610 620 630 640 650
NOV85	
		HSLYLYKDKREQTPSEEEQPISVNACLIDISYSETKRKNVFRLLTSDCE
gi	7243229	TLKFLSAHLKTVAENSEKKNMEPRNLAIIVFGPTLVRTSEDNMTHMVTHMP
gi	14736381	HSLYLYKDKREQTPSEEEQPISVNACLIDISYSETKRKNVFRLLTSDCE
gi	7959263	-----
gi	14748447	-----
gi	12856667	-----
		660 670 680 690 700
NOV85	
		CLFQAEDRDDMLAWIKTIQESSNLNEEDTGVNTRDLISRRIKEYNNLMMSK
gi	7243229	DQYKIVETLIQHHDWFFTEEGAEPLTTVQEESTVDSQPVPNIDHLLTNI
gi	14736381	CLFQAEDRDDMLAWIKTIQESSNLNEEDTGVNTRDLISRRIKEYNNLMMSK
gi	7959263	-----
gi	14748447	-----
gi	12856667	-----
		710 720 730 740 750
NOV85	
		AEQLPKTPRQSLSIQOTLLGAKSEPKTQSPHSPKEESERKLLSKDDTSP
gi	7243229	GRTGVSPGDVSDSATSDSTKSKGWSGSGKDQYSRELLVSSIFAAASRKRK
gi	14736381	AEQLPKTPRQSLSIQOTLLGAKSEPKTQSPHSPKEESERKLLSKDDTSP
gi	7959263	-----GLDDL
gi	14748447	-----
gi	12856667	-----
		760 770 780 790 800
NOV85	
		KDKGTWRKGIPSIMRKTFEKKPTATGTFGVRLDDCPPAHTNRYIPLIVDI
gi	7243229	KPKEKAQPSSEDELDNVFFKKENVEQCHNDTKEESKKESETLGRKQKII
gi	14736381	KDKGTWRKGIPSIMRKTFEKKPTATGTFGVRLDDCPPAHTNRYIPLIVDI
gi	7959263	YIGYRSYSPSFQRTTGLLHALSFRDSPFGGLPTFNLAQSPASFPPEASEP
gi	14748447	-----
gi	12856667	-----
		810 820 830 840 850
NOV85	
		CCKLVEERGLEYTGIVRPGNNAAISSMQEELNKGMAIDIDIQDDKWRDLN
gi	7243229	IAKENSTRKDPSTTKDEKISLGKESTPSEPPSPPHNSKHNSPTLSCRFA
gi	14736381	CCKLVEERGLEYTGIVRPGNNAAISSMQEELNKGMAIDIDIQDDKWRDLN
gi	7959263	PRVVRPEPSTRALEPPAEDRGDEVVLRQKPTGRKVQLTPARQMNLFQGD
gi	14748447	-----
gi	12856667	-----
		860 870 880 890 900
NOV85	
		VISSLLKSFRRKLPEPLFTNDKYADFIEANRKEDPLDRLKTLKRLIHDLP
gi	7243229	ILKESPRSLLAQKSSHLEETGSDSGTLLSTSSQASLARFSMKKSTSPETK
gi	14736381	VISSLLKSFRRKLPEPLFTNDKYADFIEANRKEDPLDRLKTLKRLIHDLP
gi	7959263	ESPEPEASGRGERLGRKVAPLATTEDSLASIPFIDEPTSPSIDLQAKHVP
gi	14748447	-----
gi	12856667	-----
		910 920 930 940 950
NOV85	
		EHHYETLKFLSAHLKTVAENSEKKNMEPRNLAIIVFGPTLVR-----T
gi	7243229	HSEFLANVSTITSYTTSSATYLTSLDSSRLSPEVQSVAESKGD---E
gi	14736381	EHHYETLKFLSAHLKTVAENSEK-NKMEPRNLAIIVFGPTLVR-----T
gi	7959263	ASAVVSSAMNSAPVLGTSPSSPTFTFTLGRHYSQDCSSIKAGRSSSYLLA
gi	14748447	-----MEPGNLAIVFGPTLVR-----T
gi	12856667	--EELSIAGTITSDYSTTSSITYLTSLDSSRLSPEVQSVAESKGD---E
		960 970 980 990 1000

```

NOV85 SEDNMTHMVT-----HMPDQYKIVETLQHHDWF---FTEEG-AEEPL
gi|7243229| ADDERSELISEGRPVETDSESEFPVFPALTTSERLFRGKLQEVTKSSRRN
gi|14736381| SEDNMTHMVT-----HMPDQYKIVETLQHHDWF---FTEEG-AEEPL
gi|7959263| ITTERSKSCDDGLNTRFDEGRVLRRLPNRIPSLRMLRSFDTGSLDSWGT
gi|14748447| SEDNMTHMVT-----HMPDQYKIVETLQHHDWF---FTEEG-AEEPL
gi|12856667| ADDERSELISEGRPLETDSESEFPVFPITLTSDRLEFRGKFOEVARVSRN

      1010      1020      1030      1040      1050
NOV85 TTVQEEESTVDSQPVPNIDHLLTNIGRTGVSFGDVSDSATSDSTKSKG---
gi|7243229| SEGSELSCTEGSLTSSLDSTRQLFSSHKLIECDTLRKKKSARFKSD---
gi|14736381| TTVQEEESTVDSQPVPNIDHLLTNIGRTGVSFGDVSDSATSDSTKSKG---
gi|7959263| SEDADAPSKRHSTSDLSDATFSDIRREGWLYYKQILTKKGGKAGSGLRQW
gi|14748447| TTVQEEESTVDSQPVPNIDHLLTNIGRTGVSFGDVSDSATSDSTKSKG---
gi|12856667| SEGSEASCTEGSLTPSLDSTRQOFSSHRLIECDTLRKKKSARFKSD---

      1060      1070      1080      1090      1100
NOV85 -----SWGSGKDQYSRELLVSSIFAAASRRKKPKKEKAQPSSSEDEL
gi|7243229| -----SGSLGDAKNEKEAPSLTKVFDVMKKCKSTGSLLTPTRGSESEK
gi|14736381| -----SWGSGKDQYSRELLVSSIFAAASRRKKPKKEKAQPSSSEDEL
gi|7959263| KRKYAALRARSLSLSKEREPGPAAGAAAAAGAGEDEAAPVCIGSCLVDI
gi|14748447| -----SWGSGKDQYSRELLVSSIFAAASRRKKPKKEKAQPSSSEDEL
gi|12856667| -----SGSEGDTRTEKETPALAKMFDVMKKCKSTGSLLTPTSRSESEK

      1110      1120      1130      1140      1150
NOV85 DNVFEEKENVEQCHNDTKESKKKESITLGRKQKIIIAKENSTRK-----
gi|7243229| QEPTWTKIADRLKLRPRAPADDMFGVGNHKNVNAETAKRKSIRRRHTLGG
gi|14736381| DNVFEEKENVEQCHNDTKESKKKESITLGRKQKIIIAKENSTRK-----
gi|7959263| SYSETKRRHVFRLLTADFCEYLFQADRDDMLGWIRAIRENSRAE---GE
gi|14748447| DNVFEEKENVEQCHNDTKESKKKESITLGRKQKIIIAKENSTRK-----
gi|12856667| QEATWTKIADRLKLRPRAPADDMFGVGNQKPTAETAKRKNIKRRHTLGG

      1160      1170      1180      1190      1200
NOV85 DPSTTKDEKISLGKESTPSEEPSPPHNSKHNSPTLSR-FAILKESPRS
gi|7243229| HRDATEISVLNFWKVEQSCERESELSAVNRLKPKCSAODLSISDWLARE
gi|14736381| DPSTTKDEKISLGKESTPSEEPSPPHNSKHNSPTLSR-FAILKESPRS
gi|7959263| DPGCANQALISKKLNDYRKVSHSSGPKADSSPKGSRGLGGLKSEFLKQSA
gi|14748447| DPSTTKDEKISLGKESTPSEEPSPPHNSEHNSPTLSR-FAILKDSPRS
gi|12856667| HRDATEISVLNFWKVEQSADESELSAVNRLKPKCSAODLSISDWLARE

      1210      1220      1230      1240      1250
NOV85 LLAQKSSHL-----EETGSDSGT-LLSTSSCASLARFS--MKKSTS
gi|7243229| RLRTSTDLRGEIGDPQTEPNPSTR---ELATDTPLSLHCNTGSSSSTL
gi|14736381| LLAQKSSHL-----EETGSDSGT-LLSTSSCASLARFS--MKKSTS
gi|7959263| ARGRLTODLPAGSKDDSAAPKTPWCINILKKKKAPRAFGVRLEECOP
gi|14748447| LLAQKSSHL-----EETGSDSGT-LLSTSSCASLARFS--MKKSTS
gi|12856667| RLRTSADLSRGEGLPQAESPSVLGT-PYSTHSPPSQOPEARVAATSTL

      1260      1270      1280      1290      1300
NOV85 PETKHSEFLANVSTITSDYSTSSATYLTSLDSSRLSPEVQSVAESKGDE
gi|7243229| ASTNRKP-----
gi|14736381| PETKHSEFLANVSTITSDYSTSSATYLTSLDSSRLSPEVQSVAESKGDE
gi|7959263| ATENORVPLIVAACCRIVEARG-----LESTGIYRVPGNNAVSSSLQEQ
gi|14748447| PETKQ-----
gi|12856667| ASTSQS-----

      1310      1320      1330      1340      1350
NOV85 ADDERSELISEGRPVETDSESEFPVFPALTTSERLFRGKLQEVTKSSRRN
gi|7243229| -----LLSIP-----PQSP
gi|14736381| ADDERSELISEGRPVETDSESEFPVFPALTTSERLFRGKLQEVTKSSRRN
gi|7959263| NRGPGDINLQDERWQDLNVISSLKSFFRKLPEPLFTDDKYNDFTIENRI
gi|14748447| -----RVFG-----QR

```

gi 12856667	-----PLFTP-----PQSP
	1360 1370 1380 1390 1400
NOV85	SEG---SELSCTEGSLTSGLDSRRQLFSSHKLECDTLRKKKSARFKSDS
gi 7243229	D-----QINGESFQNVSKNASSAANAQPHKLSETPG-----
gi 14736381	SEG---SELSCTEGSLTSGLDSRRQLFSSHKLECDTLRKKKSARFKSDS
gi 7959263	EDARERMRTLRLKLRDLPCHYYETLKLVLGHLEKTIADHSEKNKMEPRNLA
gi 14748447	-----Q-----HHHLR--LF--HHFVCC-----
gi 12856667	D-----QINRESFQNMSONASSTANIHPHKQSESPD-----
	1410 1420 1430 1440 1450
NOV85	GSLGDAKNEKEAPSLTKVFDVMKKGKSTGSLLTPTRGESKQEPWTWTKI
gi 7243229	-----TKAEFHPCL-----
gi 14736381	GSLGDAKNEKEAPSLTKVFDVMKKGKSTGSLLTPTRGESKQEPWTWTKI
gi 7959263	LVFGPTLVRTSEDNMTDMVTHMPDRYKIVETLIQHSDFFSDEEDKGERT
gi 14748447	-----LLD-----
gi 12856667	-----TKAETPP-----
	1460 1470 1480 1490 1500
NOV85	ADRLKLRPRAPADDMFGVGNHKVNAETAKRKSIRRRHTLGHRDATEISV
gi 7243229	ADRLKLRPRAPADDMFGVGNHKVNAETAKRKSIRRRHTLGHRDATEISV
gi 14736381	ADRLKLRPRAPADDMFGVGNHKVNAETAKRKSIRRRHTLGHRDATEISV
gi 7959263	PVGDKPEQAVPNIEYLLPNIGRTVPPGDPGSADLLEI-----
gi 14748447	-----
gi 12856667	-----
	1510 1520 1530 1540 1550
NOV85	LNFWKVHEQSGERESELSAVNRLKPKCSAQDLSISDWLARERLRTSTSDL
gi 7243229	LNFWKVHEQSGERESELSAVNRLKPKCSAQDLSISDWLARERLRTSTSDL
gi 14736381	LNFWKVHEQSGERESELSAVNRLKPKCSAQDLSISDWLARERLRTSTSDL
gi 7959263	-----
gi 14748447	-----
gi 12856667	-----
	1560 1570 1580 1590 1600
NOV85	SRGEIGDPQTENPSTREIATTDTPLSLHCNTGSSSSTLASTNRPLLSIPP
gi 7243229	SRGEIGDPQTENPSTREIATTDTPLSLHCNTGSSSSTLASTNRPLLSIPP
gi 14736381	SRGEIGDPQTENPSTREIATTDTPLSLHCNTGSSSSTLASTNRPLLSIPP
gi 7959263	-----
gi 14748447	-----
gi 12856667	-----
	1610 1620 1630 1640
NOV85	QSPDQINGESFQNVSKNASSAANAQPHKLSETPGKAEFHPCL
gi 7243229	QSPDQINGESFQNVSKNASSAANAQPHKLSETPGKAEFHPCL
gi 14736381	QSPDQINGESFQNVSKNASSAANAQPHKLSETPGKAEFHPCL
gi 7959263	-----
gi 14748447	-----
gi 12856667	-----

Table 85F lists the domain description from DOMAIN analysis results against NOV85. This indicates that the NOV85 sequence has properties similar to those of other proteins known to contain this domain.

Table 85F. Domain Analysis of NOV85

gnl Smart smart00324, RhoGAP, GTPase-activator protein for Rho-like GTPases; GTPase activator proteins towards Rho/Rac/Cdc42-like small GTPases. SEQ ID NO:878					
CD-Length = 175 residues, 100.0% aligned Score = 169 bits (428), Expect = 9e-43					
NOV85:	555	RYIPLIVDICCKLVEERGLEYTGIYRVPGNNAAISSMQEELNKGMAIDIDIQDDKWRDLNV	614		
		+ IP+IV+ C + +E+RGL+ GIYR G+ + + ++E + G D D++			
Sbjct:	1	KPIPIIVEKCIIEYLEKRGLDTEGIYRKSGSASRVKELREAFDSGPDPD--LDLSEYDVHD	58		
NOV85:	615	ISSLLKSFFRKLPEPLFTNDKYADFIENRKEDPLDRLKTLKRLIHDLPHEHHYETLKFLS	674		
		++ LLK F R+LPEPL T + Y +FIEA + ED +RL+ L+ L+ LP + TL++L			
Sbjct:	59	VAGLLKLFLRELPEPLITFELYEEFIEAAKLEDEEERLRALRELLSLLPPANRATLRYLL	118		
NOV85:	675	AHLKTVAENSEKKNKMEPRNLAIVFGPTLVRTSEDNMTHMVTHMPDQYKIVETLIQHHD	733		
		AHL VAE+SE +NKM RNLAIVFGPTL+R + + + Q K+VE LI++ D			
Sbjct:	119	AHLNRVAEHSE-ENKMTARNLAIVFGPTLLRPPDGEA-SLKDIRHQNKVVEFLIENAD	175		

Rho GTPases control a variety of cellular processes. There are three subtypes of Rho GTPases in the Ras superfamily of small G proteins: RHO, RAC, and CDC42. GTPase-activating proteins (GAPs) bind activated forms of Rho GTPases and stimulate GTP hydrolysis. Through this catalytic function, Rho GAPs negatively regulate Rho-mediated signals. GAPs may also serve as effector molecules and play a role in signaling downstream of Rho and other Ras-like GTPases.

By screening a Jurkat cDNA library using a yeast 2-hybrid system with an activated form of RAC as bait, followed by screening a placenta cDNA library, Toure et al. (1998) isolated a cDNA encoding RACGAP1, which they called MGCRCACGAP. The predicted 527-amino acid RACGAP1 protein has a large N-terminal region containing a protein kinase C-like cysteine-rich motif. RACGAP1 shares highest homology with the Drosophila RnRacGAP and the chimerins of rat and human. Functional analysis showed that the GAP domain of RACGAP1 exhibits strong GAP activity towards CDC42, RAC1, and RAC2. Northern blot analysis detected an approximately 3.2-kb RACGAP1 transcript that was most abundantly expressed in testis, with low expression in most other tissues. Western blot analysis detected a RACGAP1 protein of 58 kD in testis extracts. In situ hybridization showed that RACGAP1 expression is restricted to germ cells in mature testis. Human breakpoint cluster region (bcr) gene product is a member of a group of GTPase-activating proteins that act exclusively on members of the Ras-related Rho subfamily. A complementary DNA was isolated from *Caenorhabditis elegans* that encoded a polypeptide of 1438 amino acid residues, CeGAP, which contains a domain with sequence similarity to the COOH-terminal segment (GTPase-activating protein region) of Bcr and other known GTPase-activating proteins of the Rho subfamily. It also contains a "pleckstrin homology" motif, present in many signaling proteins

including GTPase-activating proteins and nucleotide exchange factors. The Bcr-like domain of CeGAP exhibited activity not only on members of the *C. elegans* and human Rho subfamily but surprisingly also on *C. elegans* Ras protein (let-60), human Ras, and Rab3A.

5 CeGAP is therefore the first GTPase-activating protein acting on Ras-related proteins across different subfamilies. Together with the presence of the pleckstrin homology motif, this suggests a central and integrative role for CeGAP in a signaling pathway common to Ras and related proteins.

NOV85 is predicted to be expressed in at least the following tissues: pancreas, stomach, brain, bone. This information was derived by determining the tissue sources of the
10 sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV85 is provided in Example 2.

The NOV85 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Alzheimer's
15 disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hypercalcaemia, ulcers, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity as well as other diseases, disorders and conditions. NOV85 nucleic acids encoding the CeGAP-
20 like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a GTPase activating protein-like protein includes the nucleic acid whose sequence is provided in Table 85A, or a fragment
25 thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 85A while still encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 85A, including nucleic acid fragments that are
30 complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least

in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

5 The novel protein of the invention includes the GTPase activating protein-like protein whose sequence is provided in Table 85B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 85B while still encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein,
10 up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section
15 below.

NOV86

The disclosed NOV86 (alternatively referred to herein as CG56957-01) includes the 3451 nucleotide sequence (SEQ ID NO:279) shown in Table 86A. A NOV86 ORF begins
20 with a Kozak consensus ATG initiation codon at nucleotides 35-37 and ends with a stop codon at nucleotides 3442-3445.

Table 86A. NOV86 Nucleotide Sequence (SEQ ID NO:279)

```

CCCGCCCCCTCGGGCTCCCGGCCGGGGCCCCCATCATGTTCTCCAGGAAGAAACGAGAGCTCATGAAAACC
CCTTCCATCTCGAAAAAGAACCGCGCGGGAAGCCCCAGCCCGCAGCCCTCGGGGGAGCTGCCCAGGAAGG
ATGGGGCTGACGCGGTGTTCCTCCGGACCAAGCCTGGAGCCGCCCGCTGGGTCTCCGGCGTCAAGGCCAC
AGGGACCTCAAGCGGCCACCAGCCTGAGCCGCCACGCCAGCGCGGCTGGCTTCCCCCTGTGCGGTGCT
GCCTCCTGGACACTGGGCCGGAGCCACCGAGCCCACTGACAGCCGCCAGCCCGGGCGAGCTGCCCAACCG
AGGGTGCCCGGCCGGACGTCGTGAGGACATCTCCCATCTGCTGGCGGACGTGGCCCGCTTCGCTGAGGG
CCTTGAGAACTTAAGGAGTGTGTGTGCGTGACGACCTCCTGAGGCCCGCCGCCCGCGGCCACGAG
TGCTGGGTGAGGCTCTGCGTGTGATGCATCAGATCATCTCCAAGTACCCGCTGCTGAACACCGTGGAGA
CGCTCACCGCAGCCGGCACCCCTCATTGCCAAGGTCAAAGCCTTCCATTATGAGAGCAACAATGATCTGGA
GAAACAGGAGTTCGAGAAGGCCCTGGAGACGATTGCTGTGGCCTTCAGTAGCACAGTGTCCGAGTTCCCTC
ATGGGTGAAGTGGACAGCAGCACCCCTCCTAGCAGTGCCTCCTGGGGACTCGAGCCAGTCCATGGAAAGCC
TGTATGAGACCGGCAGTGAGGGCACGCCCTCCAGCCTGGATGACTGTGACGCCGGCTGCCTGCCCGCGA
GGAGGTGGACGTGCTGCTACAGCGCTGTGAGGGGGCGCTGGATGCCGCACTGCTGTATGCCAAGACATG
GCCAAGTACATGAAGGACCTCATCAGCTACCTGGAGAAGCGGACGACGCTGGAGATGGAGTTTGCCAAGG
GCCTGCAGAAGATCGCTCACAACCTGCAGACAGAGCGTCATGCAGGAGCCCCACATGCCGCTCCTGTCCAT
CTACTCGCTGGCCCTGGAGCAGGACCTGGAGTTCGGCCACAGCATGGTGACAGCGGTGGGCACCTTGCA
ACCGAGACCTTCATGCAGCCCTGACCCTGCGGCGGCTTGAACACGAGAAGCGCAGGAAGGAGATCAAGG
AGGCCTGGCACCGTGCCAGAGGAAGCTGCAAGAGGCGGAGTCCAACCTGCGCAAGGCCAAGCAGGGTTA
CGTGACGCGCTGCGAGGACCACGACAAGGCTCGCTTCCTCGTGGCCAAGGCGGAGGAGGAGCAGGCTGGC
AGCGCGCCGGGAGCAGGCGGCACGGCCACCAAGACCCTGGACAAGCGCGCGGCTGGAGGAGGAGGCCA
AGAACAAGGCGGAGGAAGCTATGGCCACCTACCGCACCTGCGTGGCCGACGCGAAGACGCAGAAGCAGGA
GCTGGAGGATACCAAGGTGACGGCGCTGCGGCAGATCCAGGAGGTATCCGGCAGAGCGACCAACCATC
AAGTCGGCCACGATCTCTACTACCAGATGATGCATATGCAGACGGCGCGCTGCCCGTGCACCTCCAGA

```

TGCTGTGTGAGAGCAGCAAGCTGTATGACCCAGGCCAGCAGTACGCCTCCCACGTGCGCCAGCTGCAGCG
GGACCAGGAGCCCGATGTGCACTACGACTTTGAGCCCCACGTCTCCGCCAACGCCTGGTCCCCCGTCATG
CGTGCCCGGAAGAGCAGCTTCAACGTGAGTGATGTGGCGCGGCCGAGGCTGCCGGGAGCCCCCAGAAG
AAGGCGGGTGCACTGAGGGCACACCTGCCAAGGACACAGGGCCGGGCGAGGACACAGGTTTCAAGTC
ATGGCCGCTCTCGATCTCAGACTCGGACAGTGGGCTGGACCCCGGCCCTGGCGCAGGGGACTTTAAGAAG
TTCGAGCGGACGTCCAGTGCTACCATGTCTCGTCCAGGAGAGCTGGTGGACCCAGACGGTGGAGCCG
GGGCTTCAGCCTTTGAGCAGGCTGACCTCAACGGCATGACCCCGAGCTGCCGCTGGCCGTGCCAGTGG
ACCGTTCCGCCACGAGGGGCTGTCCAAGGCGGCCGTACTACCGGCTCCGGAAGCTCCGCACGCCCCGCC
AAGTGCCGCGAGTGCAACAGCTACGTCTACTTCCAGGGTGCTGAGTGTGAAGAGTGTGCTGGCCTGCC
ACAAGAAATGTCTGGAGACGCTGGCCATACAGTGCGGGCACAGAAGCTGCAGGGCCGCTGCAGCTGTT
CGGCCAGGACTTCAGCCACGCGGCCCGCAGCGCCCCGACGGCGTGCCCTTCATCGTCAAGAAGTGCCTG
TGCGAGATCGAGCGCGGGCGCTGCGCACCAAGGGCATCTACCGGGTCAATGGGGTAAAGACACGCGTGG
AGAAGCTGTGCCAGGCTTCGAGAACGGCAAGGAGCTGGTTCGAGCTGTGCGAGGCTTCGCCCCACGACAT
CAGCAACGTCTCAAGCTCTACCTGCGTCAGCTTCCCGAGCCGCTCATCTCTTCCGCTCTACCCAGAG
CTCGTAGGGCTGGCCAAGGACAGCCTGAAGGCAGAGGCCGAGGCCAAGGCGCGTCCCGGGGCCGCGAGG
ACGGCTCGGAGAGCGAGGCAGTGGCGGTGGCCCTGGCAGGTGCGTTCGCGGAGCTCTCGCGGACCTGCC
GCCTGAGAACCGGGCTCGCTGCGAGTACCTGCTGCGTACCTACGAGGATCGTGGAGGTGGAGCAGGAC
ACAAGATGACCCCGGAACCTGGGCATCGTGTTCGGGCCACGCTGCTTCGGGCCACGGGCCACCGAGG
CCACCGTGTCCCTCTCTCTCTGGTGGATTATCCCCATCAGGCCCGCGTCTCGAGACTCTCATCGTCCA
CTACGGCTGGTCTTCGAGGAGGAGCCGAGGAGACCCCGGGGGCCAGGACGAGTCTCAACACGCGA
GCTGAGGTAGTCTCCAGGTGCCGTACCTGGAGGCGGGCGAGGCGGTGCTTACCCGTCAGGAGGCGG
CGCGGACGGGTGCAGAGAATCCCGAGTTGTGTCCAACGATTTCGAGCTCGGACCTAGAGGAGGCTCCGA
GCTGCTGTCTCATCGAGGCCAGTGCCTGGGCCACCTCAGCTTCTCGAGCAGCAGCAGAGCGAGGCC
AGCCTAGAGGTGGCTTCTGGCAGCCACAGCGGCAGTGAGGAGCAGCTGGAGGCCACAGCCCGGAGGACG
GGGACGGGACGAGGACGGCCCGCCAGCAGCTCTCAGGATTCAACACCAACAGCTCCACAACCTGCT
GCAGGCCCCACTGCCCCCATGAGGCTCCGTGGCGGGCGATGACACTGGGCTCTGCAGGGAAGGCAG
CCGAATTCTGTGAGCTGG

A NOV86 polypeptide (SEQ ID NO:280) encoded by SEQ ID NO:279 is 1136 amino acids in length and is presented using the one-letter amino acid code in Table 86B. The Psort profile for NOV86 predicts that this sequence is likely to be localized to the nucleus with a certainty of 0.9800.

Table 86B. NOV86 Polypeptide Sequence (SEQ ID NO:280)

MFSRKKRELMTFSPISKKNRAGSPSPQPSGELPRKDGADAVFPGPSLEPPAGSSGVKATG
TLKRPTSLSRHASAAGFPLSGAASWTLGRSHRSPLTAASPGELPTEGAGPDVVEDISHLL
ADVARFAEGLEKLKECVLRDDLEARRPRAHECLGEALRVMHQIIISKYPLINTVETLTA
GTLIAKVKAHYESNNLEKQEFKALETIAVAFSSTVSEFLMGEVDSSTLLAVPPGDSS
QSMESLYGPGSEGTPPSLDDCAGCLPAEEVDVLLQRCGGVDAALLYAKNMAKYMKDLI
SYLEKRTTLEMEFAKGLQKIAHNCRQSVMQEPHMLLSIYSLALEQDLEFGHSMVQAVGT
LQTQTFMQPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLKAKQGYVQRCEDHDKAR
FLVAKAEQAGSAPGAGGTATKTLDKRRRLEEEAKNKAEEAMATYRTCVADAKTQKQEL
EDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQTAPLPVHFQMLCESSKLYDPGQY
ASHVRQLQRDQEPDVHYDFEPHVSANAWSPVMARKSSFNVSVDVARPEAAGSPPEGGCT
EGTPAKDHRAGRGHQVHKSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVD
PDGGAGASAFEQADLNGMTPELPVAVPSPGFRHEGLSKAARTHRRLKLRTPAKCRECNSY
VYFQGAECCECLACHKKCLETLAIQCGHKKLQGRQLQFLGQDFSHAARSAPDGVPIVKK
CVCEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKELVELSQASPHDISNVCLKYLRQL
PEPLISFRLYHELVLGAKDSLKAEEAKAASRGRQDGESEAVAVAGRLRELLRDLPP
ENRASLQYLLRHLRRIVEVEQDNKMTPGNLGIVFGPTLLRPRPTEATVSLSSLVDPHQ
RVETLIVHYGLVFEEPEETPGGQDESSNQRAEVVVQVPYLEAGEAVVYPLQEAADGC
RESRVVNSDSDLEEASELLSSSEASALGHLFSLEQQQSEASLEVASGSHSGSEQLEA
TAREGDGDEDEGPAQQLSGFNTNQSNNVLQAPLPPMRLRGGRTLGSCRERQPEFV

A BLAST analysis of NOV86 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV86 had high homology to other proteins as shown in Table 86C.

Table 86C. BLASTX results from PatP database for NOV86

	High Score	Smallest Sum Probability P (N)
Sequences producing High-scoring Segment Pairs:		
patp:AAU17313 Novel signal transduction pathway protein	2855	5.7e-299
patp:AAW75995 GTPase activating protein (GAP), PARG	1497	2.9e-153
patp:AAV90268 Human GTP-ase activating polypeptide PARG	1497	2.9e-153
patp:AAU17459 Novel signal transduction pathway protein	1265	1.1e-128
patp:AAU17460 Novel signal transduction pathway protein	1127	4.7e-114

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 3447 of 3451 bases (99%) identical to a gb:GENBANK-
ID:D86976|acc:D86976.1 mRNA from *Homo sapiens* (Human mRNA for KIAA0223 gene).
The full amino acid sequence of the protein of the invention was found to have 1134 of 1136
amino acid residues (99%) identical to, and 1135 of 1136 amino acid residues (99%) similar
to, the 1165 amino acid residue ptrn:SPTREMBL-ACC:Q92619 protein from *Homo sapiens*
(Human) (MYELOBLAST KIAA0223). NOV86 also has homology to the other proteins
shown in the BLASTP data in Table 86D.

Table 86D. NOV86 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 14765644 ref XP_037574.1 (XM_037574)	minor histocompatibility antigen HA-1 [<i>Homo sapiens</i>]	1136	1134/1136 (99)	1135/1136 (99)	0.0
gi 1504026 dbj BAA13212.1 (D86976)	similar to C.elegans protein (Z37093) [<i>Homo sapiens</i>]	1165	1134/1136 (99)	1135/1136 (99)	0.0
gi 2896796 gb AAC03237.1 (AC004151)	D1013901 [<i>Homo sapiens</i>]	996	994/996 (99)	995/996 (99)	0.0
gi 12857707 dbj BAB31085.1 (AK018130)	MYELOBLAST KIAA0223 (FRAGMENT)~putative [<i>Mus musculus</i>]	523	423/542 (78)	460/542 (84)	0.0
gi 13635768 ref XP_017232.1 (XM_017232)	similar to PTPL1-associated RhoGAP 1 (H. sapiens) [<i>Homo sapiens</i>]	1261	345/881 (39)	509/881 (57)	e-160

15 This BLASTP data is displayed graphically in the ClustalW in Table 86E. A multiple sequence alignment is given, with the NOV86 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 86D.

Table 86E. ClustalW Alignment of NOV86

NOV86	(SEQ ID NO:280)
gi 14765644	(SEQ ID NO:729)
gi 1504026	(SEQ ID NO:730)
gi 2896796	(SEQ ID NO:731)
gi 12857707	(SEQ ID NO:732)
gi 13635768	(SEQ ID NO:733)

	10	20	30	40	50
NOV86
gi 14765644	-----	-----	-----	-----	-----
gi 1504026	QCGPEAASAGSCSAETPSPPPRAPGRGPIMFSRKKREL	MKT	PSIS	SKKNRA	
gi 2896796	-----	-----	-----	-----	-----
gi 12857707	-----	-----	-----	-----	-----
gi 13635768	-----	-----	-----	-----	-----

	60	70	80	90	100
NOV86
gi 14765644	GSPSPQPSGELPRKDCADAVFPGPSLE	EPAGSSGVKATGTLKRPTSLSRH			
gi 1504026	GSPSPQPSGELPRKDCADAVFPGPSLE	EPAGSSGVKATGTLKRPTSLSRH			
gi 2896796	GSPSPQPSGELPRKDCADAVFPGPSLE	EPAGSSGVKATGTLKRPTSLSRH			
gi 12857707	-----	-----	-----	-----	-----
gi 13635768	WASGQLSTDITTSEMGLKSLSSNSIFDE	-----	-----	-----	-----

	110	120	130	140	150
NOV86
gi 14765644	ASAAGFPLSGAASWTLGRSHRSPLTAASPGELPTEGAGPDVVEDISHLLA				
gi 1504026	ASAAGFPLSGAASWTLGRSHRSPLTAASPGELPTEGAGPDVVEDISHLLA				
gi 2896796	ASAAGFPLSGAASWTLGRSHRSPLTAASPGELPTEGAGPDVVEDISHLLA				
gi 12857707	-----	-----	-----	-----	-----
gi 13635768	-----	-----	-----	-----	-----

	160	170	180	190	200
NOV86
gi 14765644	DVARFAEGLEKLKECVLRD	DLLEARRPRAHECLGEALRVMHQIISKYPLL			
gi 1504026	DVARFAEGLEKLKECVLRD	DLLEARRPRAHECLGEALRVMHQIISKYPLL			
gi 2896796	DVARFAEGLEKLKECVLRD	DLLEARRPRAHECLGEALRVMHQIISKYPLL			
gi 12857707	-----	-----	-----	-----	-----
gi 13635768	DIRKFSHMLLYLKEAIFFS	DCFKFVIHIRLEELRLVLKSLMNKHQNE			

	210	220	230	240	250
NOV86
gi 14765644	NTVETLTAAGTLIAKVKAFFHYESNNDLEK	QEF	EKALETIAVAFSST		
gi 1504026	NTVETLTAAGTLIAKVKAFFHYESNNDLEK	QEF	EKALETIAVAFSST		
gi 2896796	NTVETLTAAGTLIAKVKAFFHYESNNDLEK	QEF	EKALETIAVAFSST		
gi 12857707	NTVETLTAAGTLIAKVKAFFHYESNNDLEK	QEF	EKALETIAVAFSST		
gi 13635768	NSVDLQNAAEMLTAKVKAFFHYESNNDLEK	QEF	EKALETIAVAFSST		

	260	270	280	290	300
NOV86
gi 14765644	VSEFLMGEVDSSTLLAVPPGDSSQSMESLYGPGSEGTPPSLED	CD	DAGCLP		
gi 1504026	VSEFLMGEVDSSTLLAVPPGDSSQSMESLYGPGSEGTPPSLED	CD	DAGCLP		
gi 2896796	VSEFLMGEVDSSTLLAVPPGDSSQSMESLYGPGSEGTPPSLED	CD	DAGCLP		
gi 12857707	VSEFLMGEVDSSTLLAVPPGDSSQSMESLYGPGSEGTPPSLED	CD	DAGCLP		
gi 13635768	VSEFLMGEVDSSTLLAVPPGDSSQSMESLYGPGSEGTPPSLED	CD	DAGCLP		

	310	320	330	340	350
NOV86
gi 14765644	AEVDVLLQRCGGVDAALLYAKNMAKYMKD	LISYLEKRTTLEMEFAKGL			

gi	1504026	AAEVDVLLQRCCEGGVDAALLYAKNMAKYMKDLSISYLEKRTTLEMEFAKGL
gi	2896796	AAEVDVLLQRCCEGGVDAALLYAKNMAKYMKDLSISYLEKRTTLEMEFAKGL
gi	12857707	PIELDNLVLLKNTDSIELALSIAKTWSKYTKNIIVSWVEKKLNLELESTRNM
gi	13635768	PIELDNLVLLKNTDSIELALSIAKTWSKYTKNIIVSWVEKKLNLELESTRNM
		360 370 380 390 400
NOV86		QKIAHNCQSVMQEPHMPLLSIYSLALEQDLEFGHSMVQAVGTLOTQTFM
gi	14765644	QKIAHNCQSVMQEPHMPLLSIYSLALEQDLEFGHSMVQAVGTLOTQTFM
gi	1504026	QKIAHNCQSVMQEPHMPLLSIYSLALEQDLEFGHSMVQAVGTLOTQTFM
gi	2896796	QKIAHNCQSVMQEPHMPLLSIYSLALEQDLEFGHSMVQAVGTLOTQTFM
gi	12857707	QKIAHNCQSVMQEPHMPLLSIYSLALEQDLEFGHSMVQAVGTLOTQTFM
gi	13635768	VKLAEATRINTIGIQEPMPLQSLFTNALINDIESSELLOCTTAALQANKFV
		410 420 430 440 450
NOV86		QPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCEDHD
gi	14765644	QPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCEDHD
gi	1504026	QPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCEDHD
gi	2896796	QPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCEDHD
gi	12857707	QPLTLRRLEHEKRRKEIKEAWHRAQRKLQEAESNLRKAKQGYVQRCEDHD
gi	13635768	QPLLCGRKNEMEKORKEIKEIKWQEQNKMLEAENALKKAKLLCMQRQDRYE
		460 470 480 490 500
NOV86		KARFLVAKAEEEEQAGSAPGAGCTATKTLDKRRRLEEEAKNKAEEAMATYR
gi	14765644	KARFLVAKAEEEEQAGSAPGAGCTATKTLDKRRRLEEEAKNKAEEAMATYR
gi	1504026	KARFLVAKAEEEEQAGSAPGAGCTATKTLDKRRRLEEEAKNKAEEAMATYR
gi	2896796	KARFLVAKAEEEEQAGSAPGAGCTATKTLDKRRRLEEEAKNKAEEAMATYR
gi	12857707	KARFLVAKAEEEEQAGSAPGAGCTATKTLDKRRRLEEEAKNKAEEAMATYR
gi	13635768	KAKSSMFRAEEHLSSSGGLAKNLNKOLEKKRRLEEEALOKVEEANELYK
		510 520 530 540 550
NOV86		TCVADAKTQKQLEDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQT
gi	14765644	TCVADAKTQKQLEDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQT
gi	1504026	TCVADAKTQKQLEDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQT
gi	2896796	TCVADAKTQKQLEDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQT
gi	12857707	TCVADAKTQKQLEDTKVTALRQIQEVIRQSDQTIKSATISYYQMMHMQT
gi	13635768	VCVTNVBERNDLENTKRETIQAQLRTLVFCQDLTLKAVITVNLPHMOHLQA
		560 570 580 590 600
NOV86		APLPVHFQMLCESSKLYDPGQQYASHVRQLQRDQEPDVHYDFEPHVSANA
gi	14765644	APLPVHFQMLCESSKLYDPGQQYASHVRQLQRDQEPDVHYDFEPHVSANA
gi	1504026	APLPVHFQMLCESSKLYDPGQQYASHVRQLQRDQEPDVHYDFEPHVSANA
gi	2896796	APLPVHFQMLCESSKLYDPGQQYASHVRQLQRDQEPDVHYDFEPHVSANA
gi	12857707	APLPVHFQMLCESSKLYDPGQQYASHVRQLQRDQEPDVHYDFEPHVSANA
gi	13635768	ASLADSLQSLCDSAKLYDPGQBYSEFVKATNSTEEBKVDGNVNKHLNS-S
		610 620 630 640 650
NOV86		WSPVMRARKSSFNVSDVARPEAAGSPPEEGGCTEGTPAKDHRAGRGRGHQVH
gi	14765644	WSPVMRARKSSFNVSDVARPEAAGSPPEEGGCTEGTPAKDHRAGRGRGHQVH
gi	1504026	WSPVMRARKSSFNVSDVARPEAAGSPPEEGGCTEGTPAKDHRAGRGRGHQVH
gi	2896796	WSPVMRARKSSFNVSDVARPEAAGSPPEEGGCTEGTPAKDHRAGRGRGHQVH
gi	12857707	WSPVMRARKSSFNVSDVARPEAAGSPPEEGGCTEGTPAKDHRAGRGRGHQVH
gi	13635768	QPS-----GFGPANSLEDVVRLPDSSNKIEEDRCSNSADITGPSFI
		660 670 680 690 700
NOV86		KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	14765644	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	1504026	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	2896796	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	12857707	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	13635768	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
		710 720 730 740 750
NOV86		KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	14765644	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	1504026	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	2896796	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	12857707	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA
gi	13635768	KSWPLSISDSGLDPGPGAGDFKKFERTSSSGTMSSTEELVDPDGGAGA

NOV86	SAFEQADLNGMTPELPVAVPSGPFREHGLSKAARTHRLRLKLRTPAKCREC
gi 14765644	SAFEQADLNGMTPELPVAVPSGPFREHGLSKAARTHRLRLKLRTPAKCREC
gi 1504026	SAFEQADLNGMTPELPVAVPSGPFREHGLSKAARTHRLRLKLRTPAKCREC
gi 2896796	SAFEQADLNGMTPELPVAVPSGPFREHGLSKAARTHRLRLKLRTPAKCREC
gi 12857707	SADDSADLNGMDPELPVAMPSGPFREHGLSKAARTHRLRLKLRTPAKCREC
gi 13635768	SADDLDEREPSPSETGPNSLGTFKKTLMSKAALTHKFRKLRSPTKCRDC

	760	770	780	790	800
NOV86	NSYVYFQGAECCECCLACHKKCLETLAIQCGHKKLQGRQLQFGQDFSHAA				
gi 14765644	NSYVYFQGAECCECCLACHKKCLETLAIQCGHKKLQGRQLQFGQDFSHAA				
gi 1504026	NSYVYFQGAECCECCLACHKKCLETLAIQCGHKKLQGRQLQFGQDFSHAA				
gi 2896796	NSYVYFQGAECCECCLACHKKCLETLAIQCGHKKLQGRQLQFGQDFSHAA				
gi 12857707	NSYVYFQGAECCECCLACHKKCLETLAIQCGHKKLQGRQLQFGQDFSCAA				
gi 13635768	EGIVVFQGVCECECLLVCHKKCLENLVIICGHQKLPCKLHLFGAETQVA				

	810	820	830	840	850
NOV86	RSAPDGVPPFIVKKCVCEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKE				
gi 14765644	RSAPDGVPPFIVKKCVCEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKE				
gi 1504026	RSAPDGVPPFIVKKCVCEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKE				
gi 2896796	RSAPDGVPPFIVKKCVCEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKE				
gi 12857707	LSTPDGVPPFIVKKCVCEIERRALHTKGIYRVNGVKTRVEKLCQAFENGKE				
gi 13635768	KKEPDGTPFFIKTKCASEIENRALCLOGIYRVCCNKIKTEKLCQALENGMH				

	860	870	880	890	900
NOV86	LVELSQASPHDISNVLKLYLRQLPEPLISFRLYHELVLGLAKDSLKAEAEA				
gi 14765644	LVELSQASPHDISNVLKLYLRQLPEPLISFRLYHELVLGLAKDSLKAEAEA				
gi 1504026	LVELSQASPHDISNVLKLYLRQLPEPLISFRLYHELVLGLAKDSLKAEAEA				
gi 2896796	LVELSQASPHDISNVLKLYLRQLPEPLISFRLYHELVLGLAKDSLKAEAEA				
gi 12857707	LVELSQASPHDISNVLKLYLRQLPEPLISFRFYHELVLGLAKDSLKAEAEA				
gi 13635768	LVDISEFSSSHDICDVLKLYLRQLPEPFTLFRLYKEFIDLAKETQHVNEEQ				

	910	920	930	940	950
NOV86	KAASRGRODG--SESEAVAVALAGRLRELLRDLPPENRASLQYLLRHRLR				
gi 14765644	KAASRGRODG--SESEAVAVALAGRLRELLRDLPPENRASLQYLLRHRLR				
gi 1504026	KAASRGRODG--SESEAVAVALAGRLRELLRDLPPENRASLQYLLRHRLR				
gi 2896796	KAASRGRODG--SESEAVAVALAGRLRELLRDLPPENRASLQYLLRHRLR				
gi 12857707	KAASRGROCG--SESEATLAMVGRLEIMQDLPAENRATLLYLLKHLRR				
gi 13635768	ETKKNSLIEDKKWPNMCIEINRILLKSKDLLRQLPASNFNSLHFLIVHLKR				

	960	970	980	990	1000
NOV86	IVEVEQDNKMTPGNLGIVFGPTLLRPRPTEATVSLSSLVDYPHQARVIET				
gi 14765644	IVEVEQDNKMTPGNLGIVFGPTLLRPRPTEATVSLSSLVDYPHQARVIET				
gi 1504026	IVEVEQDNKMTPGNLGIVFGPTLLRPRPTEATVSLSSLVDYPHQARVIET				
gi 2896796	IVEVEQDNKMTPGNLGIVFGPTLLRPRPTEATVSLSSLVDYPHQARVIET				
gi 12857707	IVEVEQDNKMTPGNLGIVFGPTLLRPRPTATVSLSSLVDYPHQARVIET				
gi 13635768	VVDHAENKMNKSNLGVIFGPSLLRPRPTTAPITISSLAEYSNOARLVEF				

	1010	1020	1030	1040	1050
NOV86	LIVHYGLVFE-----				
gi 14765644	LIVHYGLVFE-----				
gi 1504026	LIVHYGLVFE-----				
gi 2896796	LIVHYGLVFE-----				
gi 12857707	LIVHYGLVFE-----				
gi 13635768	LITVMSQKHLFDGSLQPQDVMCSIGVVDQGC FPKPLLSPEERDIERSMKSLF				

	1060	1070	1080	1090	1100
NOV86	-----EEPEETPGGQDESSNQRAEVVVQVPYLEAG				
gi 14765644	-----EEPEETPGGQDESSNQRAEVVVQVPYLEAG				
gi 1504026	-----EEPEETPGGQDESSNQRAEVVVQVPYLEAG				
gi 2896796	-----EEPEETPGGQDESSNQRAEVVVQVPYLEAG				
gi 12857707	-----EEPEEAGCSQEGAS-----TQCGQLESA				

gi 13635768	FSSKEDIHTSESESKIFERATSFEESEKONALGKCDACLSDKAQLLLDQ
	1110 1120 1130 1140 1150
NOV86	-----EAVVYPLQEAAADGCRESRVVSNDSDSLDEEAS-----
gi 14765644	-----EAVVYPLQEAAADGCRESRVVSNDSDSLDEEAS-----
gi 1504026	-----EAVVYPLQEAAADGCRESRVVSNDSDSLDEEAS-----
gi 2896796	-----EAVVYPLQEAAADGCRESRVVSNDSDSLDEEAS-----
gi 12857707	-----EGIVFPLQEBADGSRRESHAASNDSDSELEEAS-----
gi 13635768	-----EAEASASQKIEDGKTPKPLSLKSDRSTNNVERHTPRTKIRPVSLPVDRLLL
	1160 1170 1180 1190 1200
NOV86	-----
gi 14765644	-----
gi 1504026	-----
gi 2896796	-----
gi 12857707	-----
gi 13635768	-----ASPPNERNGRNMGNVNLDFCKNPAFEGVNRKDAATTVCCKFNGFDQQTLL
	1210 1220 1230 1240 1250
NOV86	-----ELLSSSEASALGHLSFLEQQQSEASLEVASGSHSGSEEQ---
gi 14765644	-----ELLSSSEASALGHLSFLEQQQSEASLEVASGSHSGSEEQ---
gi 1504026	-----ELLSSSEASALGHLSFLEQQQSEASLEVASGSHSGSEEQ---
gi 2896796	-----ELLSSSEASALGHLSFLEQQQSEASLEVASGSHSGSEEQ---
gi 12857707	-----DPLSSSDASALHRLSFLEQ--TEAGLEEGPQSHSGSEEQ---
gi 13635768	-----QKIQDKQYEQNSLTAKITMIMPSALQEKGVTTSLQISGDHSINATQPSKP
	1260 1270 1280 1290 1300
NOV86	-----LEATAREDDGDGEDGPAQQLSGFNTNQSNNVLQAPLPPMRLR
gi 14765644	-----LEATAREDDGDGEDGPAQQLSGFNTNQSNNVLQAPLPPMRLR
gi 1504026	-----LEATAREDDGDGEDGPAQQLSGFNTNQSNNVLQAPLPPMRLR
gi 2896796	-----LEATAREDDGDGEDGPAQQLSGFNTNQSNNVLQAPLPPMRLR
gi 12857707	-----LEG---EDG-----APGPWLCHFNNTNQSNNTSRAPLPTMRLR
gi 13635768	-----YAEPVRSVREASERSSDSYPLAPVRAPRTLQPHWTTTFYKPHAPITSTR
	1310 1320 1330 1340 1350
NOV86	-----GGRMTLGSCRERQPEFV-----
gi 14765644	-----GGRMTLGSCRERQPEFV-----
gi 1504026	-----GGRMTLGSCRERQPEFV-----
gi 2896796	-----GGRMTLGSCRERQPEFV-----
gi 12857707	-----GGQITGCTSQERQPEFV-----
gi 13635768	-----GNEEKPASPSAAVPPGTDHDPHGLVVKSMPPDPKASACPGQATGQPKEDS
	1360 1370 1380
NOV86	-----
gi 14765644	-----
gi 1504026	-----
gi 2896796	-----
gi 12857707	-----
gi 13635768	-----EELGLPDVNPQCQRPRLLKRMQQFEDLEGEIQFV

Table 86F lists the domain description from DOMAIN analysis results against NOV86. This indicates that the NOV86 sequence has properties similar to those of other proteins known to contain this domain.

Table 86F. Domain Analysis of NOV86				
gnl Smart smart00324, RhoGAP, GTPase-activator protein for Rho-like GTPases; GTPase activator proteins				
towards Rho/Rac/Cdc42-like small GTPases SEQ ID NO: 879				
CD-Length = 175 residues, 98.3% aligned				
Score = 157 bits (396), Expect = 4e-39				
NOV86:	774	VPFIVKKVCVEIERRALRTKGIYRVNGVKTRVEKLCQAFENGKELV-ELSQA	SPHDISNV	832
Sbjct:	3	IP IIVEKCI EYLEKRG LDT EGIYRSGSASRVKELREAFD SGPD PDL DSEYDVH	VDVAGL	62
NOV86:	833	LKLYLRQLPEPLISFRLYHELVLGLAKDSLKAEAEAKAASRG	RQD GSESEAVAVALAGRLR	892
Sbjct:	63	LKL+LR+LPEPLI+F LY E + A + E E +		101
NOV86:	893	ELLRDLPPENRASLQYLLRHLRRIVEEQDNKMTPGNLGIVFGPTLLRPR	PTEATVSLSS	952
Sbjct:	102	-LLSLPPANRATLRYLLAHLNRVAEHSEENKMTARNLAIVFGPTLLRP	PDGE ---SAS	156
NOV86:	953	LVDYPHQARVIETLIVHY	970	
Sbjct:	157	LKD I R H Q N K V V E F L I E N A	174	

5

10

20

similarity to the COOH-terminal segment (GTPase-activating protein region) of Bcr and other known GTPase-activating proteins of the Rho subfamily. It also contains a "pleckstrin homology" motif, present in many signaling proteins including GTPase-activating proteins and nucleotide exchange factors. The Bcr-like domain of CeGAP exhibited activity not only on members of the *C. elegans* and human Rho subfamily but surprisingly also on *C. elegans* Ras protein (let-60), human Ras, and Rab3A.

CeGAP is therefore the first GTPase-activating protein acting on Ras-related proteins across different subfamilies. studies suggest a central and integrative role for CeGAP in a signaling pathway common to Ras and related proteins.

NOV86 is predicted to be expressed in at least the following tissues: pancreas, stomach, brain, bone. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV86 is provided in Example 2.

The NOV86 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hypercalcaemia, ulcers, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity as well as other diseases, disorders and conditions. NOV86 nucleic acids encoding the CeGAP-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a GTPase activating protein-like protein includes the nucleic acid whose sequence is provided in Table 86A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 86A while still encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 86A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications

include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

5 In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the GTPase activating protein-like protein whose sequence is provided in Table 86B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 10 86B while still encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic 15 methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV87

NOV87 includes two GTPase activating-like proteins, designated herein as NOV87a 20 and NOV87b.

NOV87a

The disclosed NOV87a (alternatively referred to herein as CG56886-01) includes the 994 nucleotide sequence (SEQ ID NO:281) shown in Table 87A. A NOV87a ORF begins 25 with a Kozak consensus ATG initiation codon at nucleotides 16-18 and ends with a TGA codon at nucleotides 982-984. The disclosed NOV87a maps to human chromosome 17.

Table 87A. NOV87a Nucleotide Sequence (SEQ ID NO:281)

GGGGGCCACTCTCTCATGGCCCCAAAGACAAATCCAGTAGGAAGAATGTGCTGGAGGTGAGTGGCGGGG TTGGGGAGAAGGAGGGAAGGGCTTAGTGATGCCTGCCAGCTCTCTGACAACTATTGACCTGGAAATGAC CTGGGGGTTTTTGACCCCTTAATTCAGCTACGGAGCCGAGATGGCTCTGAGTACCTGATCCAGCACGAC TCGGAGGCCATCATCAGCACCTGGCATAAGGCCATTGCTCAGGGCATCCAGGAGCTGGTAAGCAGAGCCC AGGGCCTCAGCGACTTGAGCAAGGTCCGGCACAAGCTCCGCAAGTTCCTCCAGAGCGGGCCCACTGCA GTCGCTGCGGGAGAAGGGCTACATCAAAGACCAGGTGTTCCGGCTGCGCGCTGGCCGCGCTGTGTGAGCGC GAGAGGAGCCGGGTGCCACGCTTCGTGCAGCAGTGATCCGCGCCGTCGAGGCCCGCGGTCTGGACATCG ACGGGCTGTACCGCATCAGTGAAACCTGGCCACCATCCAGAAGCTACGCTATAAGGTGGACCACGGTGA GGATGAGCGCCTTGACCTGGATGACGGGCGCTGGGAGGACGTCCACGTTATCACCAGAGCCCTGAAGCTC TTCCTTCGGGAGCTGCCCCGAGCCCTCTTCCCCTTCTCGCACTTCCGCCAGTTTCATTGCGGCCATCAGTG AGCAGGACCAGGCCCGGCGCAGCCGCTGTGTGCGTGACTTGGTGCGCTCGCTGCCCGCTCCCAACCACGA

CACTCTGCGGATGCTCTTCCAGCACCTCTGCCGAGGGTGATCGAGCACGGCGAGCAGAACCGCATGTCTG
 GTGCAGAGCGTGGCCATTGTGTTCGGGCCACGCTGCTGCCGCCGAGGTGGAAGAGACCAGCATGCCCA
 TGACCATGGTGTTCAGAACAGGTGGTGGAGCTCATCTGCAGCAGTGCGCGGACATCTTCCCGCCGCA
 CTGACTGCTGGCCT

5 A NOV87a polypeptide (SEQ ID NO:282) encoded by SEQ ID NO:281 is 322 amino acids in length and is presented using the one-letter amino acid code in Table 87B. The Psort profile for NOV87a predicts that this sequence is likely to be localized to the cytoplasm with a certainty of 0.6500.

Table 87B. NOV87a Polypeptide Sequence (SEQ ID NO:282)
MAPKDKSSRKNVLEVSQGVGEKEGRGLVMPASSLTITIDLEMTWGFLTLNSQLRSRDGSEY LIQHDSEAIISTWHKAI AQGIQELVSRAGLSKVRHKLKFLQRRPTLQSLREKGYI KDQVFGCALAALCERERSRVPFVQOCIRAVEARGLDIDGLYRISGNLATIQKLRKVDH GEDERLDLDDGRWEDVHVITGALKLFFRELPEPLFPFHFQFIAAISEQDQARRSRCVR DLVRSPLPAPNHDTLRMLFQHLCCRVI EHGEQNRMSVQSVAI VFGPTLLRPEVEETSMPT MVFQNVVLELILQQCADIFPPH

NOV87b

10 The disclosed NOV87b (alternatively referred to herein as CG56886-02) includes the 985 nucleotide sequence (SEQ ID NO:283) shown in Table 87C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 7-9 and ends with a TGA codon at nucleotides 973-975. The disclosed NOV87b maps to human chromosome 17.

Table 87C. NOV87b Nucleotide Sequence (SEQ ID NO:283)
GGGGGCCACTCTCTCATGGCCCCAAAGACAAATCCAGTAGGAAGAATGTGCTGGAGGTGAGTGGCGGGG TTGGGGAGAAAGGAGGGGAGGGGGCTTAGTGATGCCTGCCAGCTCTCTGACAACATTGACCTGGAATGAC CTGGGGGTTTTGACCCTTAATTCAGCTACGGAGCCGAGATGGCTCTGAGTACCTGATCCAGCACGAC TCGGAGGCCATCATCAGCACCTGGCATAAGGCCATTGCTCAGGGCATCCAGGAGCTGGTAAGCAGAGCCC AGGGCCTCAGCGACTTGAGCAAGGTCCGGCACAAGCTCCGCAAGTTCTCCAGAGGCGGCCCACTGCA GTCGCTGCCGGAGAAGGGCTACATCAAAGACCAGGTGTTCCGGCTGCCGCTGGCCGCGCTGTGTGAGCGC GAGAGGAGCCGGTGCCACGCTTCGTGCAGCAGTGTCATCCGCGCCGTCGAGGCCCGCGTCTGGACATCG ACGGCTGTACCGCATCAGTGGAACCTGGCCACCATCCAGAAGCTACGCTATAAGGTGGACCACGGTGA GGATGAGCGCCTTGACCTGGATGACGGGCGCTGGGAGGACGTCACGTTATCACCGGAGCCCTGAAGCTC TTCTTTCGGGAGCTGCCCGAGCCCCCTTCCCTTCTCGCACTTCCGCCAGTTCAATTGCGGCCATCAGTG AGCAGGACCAGGCCCGGCGCAGCCGCTGTGTGCGTGACTTGGTGCGCTCGCTGCCCCGCTCCCAACCACGA CACTCTGCGGATGCTCTTCCAGCACCTCTGCCGAGGGTGATCGAGCACGGCGAGCAGAACCGCATGTCTG GTGCAGAGCGTGGCCATTGTGTTCCGGGCCACGCTGCTGCCGCCGAGGTGGAAGAGACCAGCATGCCCA TGACCATGGTGTTCAGAACAGGTGGTGGAGCTCATCTGCAGCAGTGCGCGGACATCTTCCCGCCGCA CTGACTGCTGGCCT

15 The NOV87b polypeptide (SEQ ID NO:284) encoded by SEQ ID NO:283 is 322 amino acids in length and is presented using the one-letter amino acid code in Table 87D. The Psort profile for NOV87b predicts that this sequence has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.6500.

Table 87D. NOV87b Polypeptide Sequence (SEQ ID NO:284)

MAPKDKSSRKNVLEVSGGVGEKEGRGLVMPASSLTITIDLEMTWGFLLNSQLRSRDGSEY
LIQHDSEAIISTWHKAI AQIQELVSRAQGLSDLKSVRHKLKFLQRRPTLQSLREKGYI
KDQVFGCALAALCERERSRVPRFVQQCIRAVEARGLDIDGLYRISGNLATIQKLRKVDH
GEDERLDLDDGRWEDVHVITGALKLFFRELPEPLFFFSHFRQFIAAISEQDQARRSRCVR
DLVRS L P A P N H D T L R M L F Q H L C R R V I E H G E Q N R M S V Q S V A I V F G P T L L R P E V E E T S M P M T
M V F Q N Q V V E L I L Q Q C A D I F P P H

A BLAST analysis of NOV87 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV87 had high
5 homology to other proteins as shown in Table 87E.

Table 87E. BLASTX results from PatP database for NOV87

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)
patp:AAU17449 Novel signal transduction pathway protein	827	2.9e-82
patp:AAB68548 Human GTP-binding associated protein	779	3.5e-77
patp:AAG66505 GTP enzyme Rho family active site 90	779	3.5e-77
patp:AAB64387 Amino acid sequence of human intracellular	711	5.6e-70
patp:AA94450 Human inflammation associated protein	589	3.0e-62

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 276 of 417 bases (66%) identical to a gb:GENBANK-
10 ID:BC006107|acc:BC006107.1 mRNA from *Homo sapiens* (*Homo sapiens*, clone MGC:12959, mRNA). The full amino acid sequence of the protein of the invention was found to have 173 of 319 amino acid residues (54%) identical to, and 229 of 319 amino acid residues (71%) similar to, the 316 amino acid residue ptnr:SPTREMBL-ACC:Q9NT76 protein from *Homo sapiens* (Human) (HYPOTHETICAL 36.4 KDA PROTEIN). NOV87 also has
15 homology to the other proteins shown in the BLASTP data in Table 87F.

Table 87F. NOV87 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 11360091 p ir T46471	hypothetical protein DKFZp434L0130.1 - human	316	168/355 (47)	221/355 (61)	7e-80
gi 13676443 d bj BAB41146.1 (AB060206)	hypothetical protein [Macaca fascicularis]	847	164/356 (46)	220/356 (61)	6e-79

gi 14245732 d bj BAB56159.1 (AB051853)	rho-GTPase activating protein [<i>Homo sapiens</i>]	731	130/328 (39)	188/328 (56)	9e-57
gi 18146831 d bj BAB83128.1 (AB030239)	RGL1 [<i>Homo sapiens</i>]	547	130/328 (39)	188/328 (56)	2e-56
gi 15080081 g b AAH11820.1 AAH11820 (BC011820)	Unknown (protein for IMAGE:3619501) [<i>Homo sapiens</i>]	599	130/328 (39)	188/328 (56)	2e-56

This BLASTP data is displayed graphically in the ClustalW in Table 87G. A multiple sequence alignment is given, with the NOV87a and b proteins being shown on lines 1 and 2 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 87F.

Table 87G. ClustalW Alignment of NOV87	
NOV87a	(SEQ ID NO:282)
NOV87b	(SEQ ID NO:284)
gi 11360091	(SEQ ID NO:734)
gi 13676443	(SEQ ID NO:735)
gi 14245732	(SEQ ID NO:736)
gi 18146831	(SEQ ID NO:737)
gi 15080081	(SEQ ID NO:738)
<div> <div>1020304050</div> <div> <div>NOV87a</div> <div>NOV87b</div> <div>gi 11360091 </div> <div>gi 13676443 </div> <div>gi 14245732 </div> <div>gi 18146831 </div> <div>gi 15080081 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>MKMADRSGKIIPGQAYIEVEYDYEYEAKDRKIVIKQGERYILVKKTNDWD</div> <div>-----</div> <div>-----</div> <div>-----M</div> <div>-----</div> <div>-----</div> </div> </div>	
<div> <div>60708090100</div> <div> <div>NOV87a</div> <div>NOV87b</div> <div>gi 11360091 </div> <div>gi 13676443 </div> <div>gi 14245732 </div> <div>gi 18146831 </div> <div>gi 15080081 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>WQVKPDENSKAFYVPAQYVKEVTRKALMPPVKQVAGLPNNSTKIMQSLHL</div> <div>-----</div> <div>-----</div> <div>LSSRWWPSSWGILGLGPRSPPRGSQLCALYAFTYTGDGQQVSLAEGDRF</div> <div>-----</div> <div>-----</div> <div>-----</div> </div> </div>	
<div> <div>110120130140150</div> <div> <div>NOV87a</div> <div>NOV87b</div> <div>gi 11360091 </div> <div>gi 13676443 </div> <div>gi 14245732 </div> <div>gi 18146831 </div> <div>gi 15080081 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>QRSTENVNKLPELSSFGKPSVVQGTGLTRDANQNFGPSYNPGHTVNLSL</div> <div>-----</div> <div>-----</div> <div>LLLRKTNSDWLARRLEAPTSRPIFVPAAYMIEESIPSQSPTTVIPGQL</div> <div>-----</div> <div>-----</div> <div>-----</div> </div> </div>	
<div> <div>160170180190200</div> <div> <div>NOV87a</div> <div>NOV87b</div> <div>gi 11360091 </div> <div>gi 13676443 </div> </div> <div> <div>..... </div> <div>-----</div> <div>-----</div> <div>DLTHNNGKFNNDSHSPKVSSQNRTLFGHFPGPEFLDVEKTSFSQEQSCD</div> <div>-----</div> <div>-----</div> <div>-----</div> </div> </div>	

gi	14245732	LWTPGPKLFHGSLEELSQUALPSRAQASSEQPPPLPRKMCRSVSTDNLSPS
gi	18146831	-----
gi	15080081	-----PPLPRKMCRSVSTDNLSPS
		210 220 230 240 250
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	SAGEGSEIRHQDSESGDELSSSSTEQIRATTPPNQGRPDSPVYANLQELK
gi	14245732	LLKPFQEGPSGRSLSQEDLPSEAS---ASTAGPQPLMSEPPVYCNLVDLR
gi	18146831	-----MSEPPVYCNLVDLR
gi	15080081	LLKPFQEGPSGRSLSQEDLPSEAS---ASTAGPQPLMSEPPVYCNLVDLR
		260 270 280 290 300
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	ISQSALPPLPGSPAIIQINGEWETHKD-SSGRCYYYDRGTQERTWKPPRWT
gi	14245732	RCPRSPPPGPACPLLQRLDAWEQHLDPNSGRCFYINSLTGCKSWKPPRRS
gi	18146831	RCPRSPPPGPACPLLQRLDAWEQHLDPNSGRCFYINSLTGCKSWKPPRRS
gi	15080081	RCPRSPPPGPACPLLQRLDAWEQHLDPNSGRCFYINSLTGCKSWKPPRRS
		310 320 330 340 350
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	RDASISKGDFQSPGDQELLSSEENYYSTSYSQSDSQCGSPPRGWSEELDE
gi	14245732	R-----SE-----
gi	18146831	R-----SE-----
gi	15080081	R-----SE-----
		360 370 380 390 400
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	RGHTLYTSDYTNEKWLKHIDDQGRQYYSADGSRSEWELPKYNASSQQQR
gi	14245732	-----TNPGSMEGTQTLKRNNVDVLQPQA
gi	18146831	-----TNPGSMEGTQTLKRNNVDVLQPQA
gi	15080081	-----TNPGSMEGTQTLKRNNVDVLQPQA
		410 420 430 440 450
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	EIIKSRSLDRRLQEPIVLTKWRHSTIVLDTNDKESPTASKPCFPENESSP
gi	14245732	KGFRSDTGTPEPLDPQGSLSLSQRTSQLDPPALQAP---RP-----
gi	18146831	KGFRSDTGTPEPLDPQGSLSLSQRTSQLDPPALQAP---RP-----
gi	15080081	KGFRSDTGTPEPLDPQGSLSLSQRTSQLDPPALQAP---RP-----
		460 470 480 490 500
	
NOV87a		-----
NOV87b		-----
gi	11360091	-----
gi	13676443	SSPKHQDTASSPKDQEKYGLLNVTKIAENGKKVRKNWLSSWAVLQGSLL
gi	14245732	----LPQLLDDDPHEVEKSGLLNMTKIAQGGKRLRKNWGPSWVVLGTGNSLV
gi	18146831	----LPQLLDDDPHEVEKSGLLNMTKIAQGGKRLRKNWGPSWVVLGTGNSLV
gi	15080081	----LPQLLDDDPHEVEKSGLLNMTKIAQGGKRLRKNWGPSWVVLGTGNSLV
		510 520 530 540 550
	
NOV87a		-----MAP-KDKSSRKNVLEVS
NOV87b		-----MAP-KDKSSRKNVLELR
gi	11360091	-----MAS-KDKSSKKNVFELR

gi	13676443	FTKT--QGSSTSWFGSNQSKPEFTVDLKGATIEMAS-KDKSSKKNVFELK
gi	14245732	FYREPPPTAPSSGWGPAGSRPESSVDLRG--AALAHGRHLSRRNVLHIR
gi	18146831	FYREPPPTAPSSGWGPAGSRPESSVDLRG--AALAHGRHLSRRNVLHIR
gi	15080081	FYREPPPTAPAGWGPAGSRPESSVDLRG--AALAHGRHLSRRNVLHIR
		560 570 580 590 600
NOV87a		GGVCEKEGRGLVMPASSITTTIDLEMTWGFLT--NSQLRSRDGSEYLIQH
NOV87b		SRDGS-EYLIQHDSIAITSTWHKATAAGIQEL--SAELPPEESESRRVDF
gi	11360091	TROCT-ELLIQSDNDITVINDWFKVLSSTIN---NQAVTDEGIEEEEI-P
gi	13676443	TROCT-ELLIQSDNDITVINDWFKVLSSTIN---NQAVTDEGIEEEEILP
gi	14245732	TIPCH-EFLLQSDHETELRAWHRALRTVIERLDRENPLELRLSGSGPAEL
gi	18146831	TIPCH-EFLLQSDHETELRAWHRALRTVIERLDRENPLELRLSGSGPAEL
gi	15080081	TIPCH-EFLLQSDHETELRAWHRALRTVIERLDRENPLELRLSGSGPAEL
		610 620 630 640 650
NOV87a		DSEAIISTWHKATAAGIQELVSRAQG---LSDLSKVRHKLKFLQRRP
NOV87b		GSSERLGSWQEKEDARPNAAPALGPVGL-ESLSKVRHKLKFLQRRP
gi	11360091	DSPGTEKHDKKEKEQKDPKILRSFKVSSID--SSECKTKKNLKKFLTRRP
gi	13676443	DSPGTEKHDKKEKEQKDPKILRSFKVSSID--SSECKTKKNLKKFLTRRP
gi	14245732	SAGEDDEEESLVSCKPLLRSSRRSSIRGPEGTEONRVRNKLRLIAKRP
gi	18146831	SAGEDDEEESLVSCKPLLRSSRRSSIRGPEGTEONRVRNKLRLIAKRP
gi	15080081	SAGEDDEEESLVSCKPLLRSSRRSSIRGPEGTEONRVRNKLRLIAKRP
		660 670 680 690 700
NOV87a		TLOSLREKGYIKDQVFGCALAALCERERSRVPRFVQOCIRAVEARGLDID
NOV87b		TLOSLREKGYIKDQVFGCALAALCERERSRVPRFVQOCIRAVEARGLDID
gi	11360091	TLOAVREKGYIKDQVFGSNLANLCQRENGTVPKFVKLCIEHVEEHGLDID
gi	13676443	TLOAVREKGYIKDQVFGSNLANLCQRENGTVPKFVKLCIEHVEEYGLDID
gi	14245732	PLQSLQERGLLRDQVFGCOLESCLCOREGDTVPSEFLRLCIAAVDKRGLDID
gi	18146831	PLQSLQERGLLRDQVFGCOLESCLCOREGDTVPSEFLRLCIAAVDKRGLDID
gi	15080081	PLQSLQERGLLRDQVFGCOLESCLCOREGDTVPSEFLRLCIAAVDKRGLDID
		710 720 730 740 750
NOV87a		GLYRISGNLATTQKLRYKVDHG-----EDERLDDLDDGRW
NOV87b		GLYRISGNLATTQKLRYKVDH-----DERLDDLDDGRW
gi	11360091	GIYRVSGNLAVIQKLRFVAVNH-----DEKLDDLNDSKW
gi	13676443	GIYRVSGNLAVIQKLRFVAVNH-----DEKLDDLNDSKW
gi	14245732	GIYRVSGNLAVIQKLRFVAVNDRERAVTSDGRYVFPEQPGQEGRLDLSTEW
gi	18146831	GIYRVSGNLAVIQKLRFVAVNDRERAVTSDGRYVFPEQPGQEGRLDLSTEW
gi	15080081	GIYRVSGNLAVIQKLRFVAVNDRERAVTSDGRYVFPEQPGQEGRLDLSTEW
		760 770 780 790 800
NOV87a		EDVHVITGALKLFFRELPEPLFFFSHFROITAAISEQDQARRSRCVRDLV
NOV87b		EDVHVITGALKLFFRELPEPLFFFSHFROITAAIKLQDQARRSRCVRDLV
gi	11360091	EDIHVITGALKMFFRELPEPLFTFNHFNDVNAIK-QEPRORVAAVKDLI
gi	13676443	EDIHVITGALKMFFRELPEPLFTFNHFNDVNAIK-QEPRORVAAVKDLI
gi	14245732	DDIHVITGALKLFLRELPOPLVEPLLLPHERAALALSESEQCLSQIQELI
gi	18146831	DDIHVITGALKLFLRELPOPLVEPLLLPHERAALALSESEQCLSQIQELI
gi	15080081	DDIHVITGALKLFLRELPOPLVEPLLLPHERAALALSESEQCLSQIQELI
		810 820 830 840 850
NOV87a		RSIPAPNHDTLRMLFQHLCCR-VIEHGEONRMSVQSVATVFGPTLLRPEVE
NOV87b		RSIPAPNHDTLRMLFQHLCCR-VIEHGEONRMSVQSVATVFGPTLLRPEVE
gi	11360091	ROLPKPNQDTMCILFRHLRR-VIENGKNRMTYQSTATVFGPTLLKPEKE
gi	13676443	ROLPKPNQDTMCILFRHLKR-VIENGKNRMTYQSTATVFGPTLLKPEKE
gi	14245732	CSMPKPNHDTLRYLLEHLCCR-VIAHSDKNRMTPHNLGIVFGPTLLRPEQE
gi	18146831	CSMPKPNHDTLRYLLEHLCCR-VIAHSDKNRMTPHNLGIVFGPTLLRPEQE
gi	15080081	CSMPKPNHDTLRYLLEHLCCR-VIAHSDKNRMTPHNLGIVFGPTLLRPEQE
		860 870
NOV87a		ETSMPTMTVFONOVVELLLQQCADIFPPH
NOV87b		ETSMPTMTVFONOVVELLLQQCADIFPPH

gi 11360091	TGNI	AVHTVYQNOIV	ELIL	LELSSIFGR-
gi 13676443	TGNI	AVHTVYQNOIV	ELIL	LELSSIFGR-
gi 14245732	TSDP	AAHALYPGOLV	OLMLTNFTSLFP--	
gi 18146831	TSDP	AAHALYPGOLV	OLMLTNFTSLFP--	
gi 15080081	TSDP	AAHALYPGOLV	OLMLTNFTSLFP--	

Table 87H lists the domain description from DOMAIN analysis results against NOV87. This indicates that the NOV87 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 87H. Domain Analysis of NOV87				
gnl Smart smart00324, RhoGAP, GTPase-activator protein for Rho-like GTPases; GTPase activator proteins towards Rho/Rac/Cdc42-like small GTPases SEQ ID NO: 879				
CD-Length = 175 residues, 99.4% aligned Score = 144 bits (364), Expect = 6e-36				
NOV87: 138	SRVPRFVQQCIRAVEARGLDIDGLYRISGNLATIQKLR	YKVDHGEDERLDDLDDGRWEDVH	197	
	+P V++CI +E RGLD +G+YR SG+ + +++LR	D G D LDL + DVH		
Sbjct: 1	KPIPIIVEKCI	EYLEKRGLDTEGIYRKSGSASRVKELREAFDSGPD	DLSE---	YDVH 57
..				
NOV87: 198	VITGALKLFFRELPEPLFFFSHFRQFIAAISEQDQARRSRCVRDLVRS	LPAPNHDTLRML	257	
	+ G LKLF RELPEPL F + +FI A +D+ R R +R+L+ LP N TLR L			
Sbjct: 58	DVAGLLKLFLRELPEPLITFELYEEFIEAAKLEDEEERLRALRELLS	LLPPANRATLRYL	117	
.				
NOV87: 258	FQHL	CRRVIEHGEQNRMSVQSV	AI	VFGPTLLRPEVEETSM
	HL RV EH E+N+M+ +++A	I	VFGPTLLRP E++ + QN+VVE +++	
Sbjct: 118	LAHL-NRVAEHSEENKMTARNL	AI	VFGPTLLRPPDGESASLKD	IRHQNKVVEFLIENA 174

Rho GTPases control a variety of cellular processes. There are 3 subtypes of Rho GTPases in the Ras superfamily of small G proteins: RHO, RAC, and CDC42. GTPase-activating proteins (GAPs) bind activated forms of Rho GTPases and stimulate GTP hydrolysis. Through this catalytic function, Rho GAPs negatively regulate Rho-mediated signals. GAPs may also serve as effector molecules and play a role in signaling downstream of Rho and other Ras-like GTPases.

By screening a Jurkat cDNA library using a yeast 2-hybrid system with an activated form of RAC as bait, followed by screening a placenta cDNA library, Toure et al. (1998) isolated a cDNA encoding RACGAP1, which they called MGCACGAP. The predicted 527-amino acid RACGAP1 protein has a large N-terminal region containing a protein kinase C-like cysteine-rich motif. RACGAP1 shares highest homology with the Drosophila RnRacGAP and the chimerins of rat and human. Functional analysis showed that the GAP domain of RACGAP1 exhibits strong GAP activity towards CDC42, RAC1, and RAC2. Northern blot analysis detected an approximately 3.2-kb RACGAP1 transcript that was most abundantly expressed in testis, with low expression in most other tissues. Western blot analysis detected a

RACGAP1 protein of 58 kD in testis extracts. In situ hybridization showed that RACGAP1 expression is restricted to germ cells in mature testis Human breakpoint cluster region (bcr) gene product is a member of a group of GTPase-activating proteins that act exclusively on members of the Ras-related Rho subfamily.

5 A complementary DNA was isolated from *Caenorhabditis elegans* that encoded a polypeptide of 1438 amino acid residues, CeGAP, which contains a domain with sequence similarity to the COOH-terminal segment (GTPase-activating protein region) of Bcr and other known GTPase-activating proteins of the Rho subfamily. It also contains a "pleckstrin homology" motif, present in many signaling proteins including GTPase-activating proteins
10 and nucleotide exchange factors. The Bcr-like domain of CeGAP exhibited activity not only on members of the *C. elegans* and human Rho subfamily but surprisingly also on *C. elegans* Ras protein (let-60), human Ras, and Rab3A.

CeGAP is therefore the first GTPase-activating protein acting on Ras-related proteins across different subfamilies. studies suggest a central and integrative role for CeGAP in a
15 signaling pathway common to Ras and related proteins.

NOV87 is predicted to be expressed in at least the following tissues: pancreas, stomach, brain, bone. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for
20 NOV87 is provided in Example 2.

The NOV87 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,
25 leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hypercalceimia, ulcers, diabetes, Von Hippel-Lindau (VHL) syndrome, pancreatitis, obesity as well as other diseases, disorders and conditions. NOV87 nucleic acids encoding the CeGAP-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be
30 assessed.

The novel nucleic acid of the invention encoding a GTPase activating protein-like protein includes the nucleic acid whose sequence is provided in Table 87A or 96C, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 87A or 96C while still

encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 87A or 96C, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the GTPase activating protein-like protein whose sequence is provided in Table 87B or 96D. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 87B or 96D while still encoding a protein that maintains its GTPase activating protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV88

The disclosed NOV88 (alternatively referred to herein as CG56394-01) includes the 1092 nucleotide sequence (SEQ ID NO:285) shown in Table 88A. A NOV88 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 25-27 and ends with a stop codon at nucleotides 1033-1035. The disclosed NOV88 maps to human chromosome 2.

Table 88A. NOV88 Nucleotide Sequence (SEQ ID NO:285)

GCACCAGCCACATCCTGAGATACCATGGTTAAGGTGAAGGCCAGAGTCAACAGATTTGGCCACATTGGGC ACCAGATCACCAGGGCTGCTTTAACTCTGGTAAAGTGATATTGTTGCCATCAGTGACCCCTTCACTGG CCTCAACTACATGGTCTACGTGTTCCAGTGTGGTTTACCCATGGCAAATCCATGGCACTGTCAAGGCT GAGAATGGGAAGCTTGTCATTAACGGAATCTCATCACCATCTTTCAGGAGCGAGATCCACCAAATCA AATGGGACAATGTTGACGCTGAGTACATTGGGTGTCCACCGTGTCTTCACCACCACAGAGAAGGCTGG

GGCTCACTTGACGAGGGAGCCAAAAGGGTCATAATCTCTACTCCCTCTGCTGACGCCCCCATGTTTCATG
 ATGGGCGTGAACCATAAGAAATATGAAAACAGCCTCAAGATCATCAGCAATGCCTCCTGTACCACCAACT
 TCTTAGCCTCCCTGGCCAAGCTCATCCATGACAACTTGGTATGTGGGAAGGACTCATGACCACGACCCA
 CACCATCACTGCCACCCAGAAGACTGTAGATGGACCTCCAGGAACTGTGGTGTGATGGCCACGGGGCT
 CTCCAGATCATCATCCCTGCATCTACTGGTGCTGCCAAAGCTGTAGGCAAGGTCATCCCGAGATGAATG
 GGAAGATTACTAGCATGGCCTTCCGTGTCCCCACCACCAATGTGTGGTTCATGCATCTGACCTGCCATCT
 GGAAATCCTGCCAAATATGATGACATCAAGAAGGTGGTGAACAGGCATCAGAGGCCCCCTCCCTCAAG
 GGCATCCTGGACTACACTGAGCACCACGTTGTCTCTCCAGCTTTAACAGTGACACCCACTCTTCCACCT
 TCAATGATGGGGCTGGTATTGCCCTCAATGACCATT'TTGTCAAGCTCATTTCCTGTTATGACAATGCATT
 TGGCTACAACAACAGGGCAGTGGACCTCATGGCCCATGGCCTCCAAGAAGTAAGACCCCGAGACCAC
 AGCCTCAGGCCCTCAGCTGCTAGGAATCCCTATTGCACTAG

A NOV88 polypeptide (SEQ ID NO:286) encoded by SEQ ID NO:285 is 336 amino acids in length and is presented using the one-letter amino acid code in Table 88B. The Psort profile for NOV88 predicts that this sequence has no signal peptide and is likely to be localized to microbodies with a certainty of 0.4804. In alternative embodiments, a NOV88 polypeptide is located to the mitochondrial matrix space with a certainty of 0.3600.

Table 88B. NOV88 Polypeptide Sequence (SEQ ID NO:286)

MVKVKARVNRFGHHGHQITRAAFNSGKVDIVAISDPFTGLNYMVYVFQCGSTHGKFKHGTV
 KAENGKLVINGNLITIFQERDPTKIKWDNVDAEYI WVSTGVFTTTEKAGAHLQQGAKRVI
 ISTPSADAPMFMMGVNHKKYENSLKIIISNASCTTNFLASLAKLIHDFGIVEGLMTTHT
 ITATQKTVDGSPSRKLWCDGHGALQIIIPASTGAAKAVGKVIPEMNGKITSMAFRVPTTNV
 SVMHLTCHLENPAKYDDIKKVKQASEAPPLKGLDYTEHHVSVSSFSNDTHSSTFNDGA
 GIALNDHFVKLISCYDNFAFGYNNRAVDLMAHMASK

A BLAST analysis of NOV88 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV88 had high homology to other proteins as shown in Table 88C.

Table 88C. BLASTX results from PatP database for NOV88

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P(N)
patp:AA07036 Breast cancer associated antigen precursor	1407	9.9e-144
patp:AA05368 Human HCMV inducible gene protein	1407	9.9e-144
patp:AAG64817 Human G3PDH fragment - <i>Homo sapiens</i> , 327 aa.	1377	1.5e-140
patp:AAE04373 Mouse cancer associated antigen OY-MC-2	1316	4.4e-134
patp:AAR12995 GAP-DH - <i>Aspergillus oryzae</i> (ATCC 42149)	993	7.4e-100

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 928 of 1055 bases (87%) identical to a gb:GENBANK-ID:AF261085|acc:AF261085.1 mRNA from *Homo sapiens* (glyceraldehyde-3-phosphate dehydrogenase (GADPH) mRNA). The full amino acid sequence of the protein of the invention was found to have 278 of 336 amino acid residues (82%) identical to, and 294 of

336 amino acid residues (87%) similar to, the 335 amino acid residue ptnr:TREMBLNEW-ACC:AAG01996 protein from *Homo sapiens* (Human) (CLONE CDABP0047 MRNA SEQUENCE). NOV88 also has homology to the other proteins shown in the BLASTP data in Table 88D.

5

Table 88D. NOV88 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7669492 ref NP_002037.2 (NM 002046)	glyceraldehyde-3-phosphate dehydrogenase [<i>Homo sapiens</i>]	335	278/336 (82)	294/336 (86)	e-152
gi 31645 emb CA A25833.1 (X01677)	glyceraldehyde-3-phosphate dehydrogenase [<i>Homo sapiens</i>]	335	227/336 (82)	294/336 (87)	e-152
gi 2407184 gb A AB94053.1 (AF017079)	glyceraldehyde 3-phosphate dehydrogenase [<i>Sus scrofa</i>]	333	272/333 (81)	290/333 (86)	e-149
gi 6983849 dbj BAA90818.1 (AB038241)	glyceraldehyde-3-phosphate dehydrogenase [<i>Felis catus</i>]	333	267/333 (80)	285/333 (85)	e-145
gi 2506441 sp P 00355 G3P_PIG	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)	333	265/333 (79)	284/333 (84)	e-145

This BLASTP data is displayed graphically in the ClustalW in Table 88E. A multiple sequence alignment is given, with the NOV88 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 88D.

10

Table 88E. ClustalW Alignment of NOV88	
NOV88	(SEQ ID NO:286)
gi 7669492	(SEQ ID NO:739)
gi 31645	(SEQ ID NO:740)
gi 2407184	(SEQ ID NO:741)
gi 6983849	(SEQ ID NO:742)
gi 2506441	(SEQ ID NO:743)
<div> <div>1020304050</div> <div> NOV88MVKVKARVNRFGHICHQITRAAFNSGKVDIVATSDPFTGLNYMVYVFOCG gi 7669492 MGKVKVG VNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYD gi 31645 MGKVKVG VNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYD gi 2407184 --MVKVG VNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYD gi 6983849 --MVKVG VNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYD gi 2506441 --MVKVG VNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYD </div> </div>	
<div> <div>60708090100</div> <div> NOV88STHGKFHGTVKAENGKLVINGNLPITIFQERDPSKIKWGNVDAEYVWVSTG gi 7669492 STHGKFHGTVKAENGKLVINGNLPITIFQERDPSKIKWGDAGA EYVVESTG gi 31645 STHGKFHGTVKAENGKLVINGNLPITIFQERDPSKIKWGDAGA EYVVESTG gi 2407184 STHGKFHGTVKAENGKLVINGNLPITIFQERDPSKIKWGDAGA EYVVESTG </div> </div>	

gi 6983849	STHGKFHCTVKAENGLVINGKPTITIFQERDPANIKWGDAGAAYVVESTG
gi 2506441	STHGKFHGTVKAENGLVINGKATITIFQERDPANIKWGDAGATVVESTG
	110 120 130 140 150
NOV88	VFTTTEKAGAHLOGGAKRVIISTPSADAPMFVMGVNHEKYDNLKIIISNA
gi 7669492	VFTTMEKAGAHLOGGAKRVIISAPSADAPMFVMGVNHEKYDNLKIIISNA
gi 31645	VFTTMEKAGAHLOGGAKRVIISAPSADAPMFVMGVNHEKYDNLKIIISNA
gi 2407184	VFTTMEKAGAHLOGGAKRVIISAPSADAPMFVMGVNHEKYDNLKIIISNA
gi 6983849	VFTTMEKAGAHLOGGAKRVIISAPSADAPMFVMGVNHEKYDNLKIIISNA
gi 2506441	VFTTMEKAGAHLOGGAKRVIISAPSADAPMFVMGVNHEKYDNLKIIISNA
	160 170 180 190 200
NOV88	SCTTNCLASLAKLIHDNFGIVEGLMTTTHITATQKTVDGSPSRKLWCDGH
gi 7669492	SCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGSPSGKLWRDGR
gi 31645	SCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGSPSGKLWRDGR
gi 2407184	SCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGSPSGKLWRDGR
gi 6983849	SCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGSPSGKLWRDGR
gi 2506441	SCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGSPSGKLWRDGR
	210 220 230 240 250
NOV88	GALQIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
gi 7669492	GALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
gi 31645	GALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
gi 2407184	GALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
gi 6983849	GALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
gi 2506441	GALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTCRLE
	260 270 280 290 300
NOV88	NPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
gi 7669492	KPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
gi 31645	KPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
gi 2407184	KPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
gi 6983849	KPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
gi 2506441	KPAKYDDIKKVVQASEGLKGLGYTEHQVSSDFNSDTHSSTFDAGA
	310 320 330
NOV88	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE
gi 7669492	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE
gi 31645	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE
gi 2407184	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE
gi 6983849	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE
gi 2506441	GIALNDHFVKLISYDNEFGYSNRVVDLMAHMASKE

Table 88F lists the domain description from DOMAIN analysis results against NOV88. This indicates that the NOV88 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 88F. Domain Analysis of NOV88		
gnl Pfam pfam02800, gpdh_C, Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain. GAPDH is a tetrameric NAD-binding enzyme involved in glycolysis and glyconeogenesis. C-terminal domain is a mixed alpha/antiparallel beta fold. SEQ ID NO:880		
CD-Length = 163 residues, 100.0% aligned		
Score = 169 bits (427), Expect = 3e-43		
NOV88:	153 TTNFLASLAKLIHDNFGIVEGLMTTTHITATQKTVDGSPSRKLWCDGHGALQIIPASTG	212

Sbjct:	1	TTN LA LAK+++DNFGI +GLMTT H TA QK VDGP K G A IIP STG	60
		TTNCLAPLAKVLNDNFGIEKGLMTTVHAYTADQKLVDPHKKDLRRGAAAPNIPTSTG	
NOV88:	213	AAKAVGKVIPEMNGKITSMAFRVPTTNVSMHLTCHLENPAKYDDIKKVVQASEAPPLK	272
		AAKAVG V+PE+NGK+T MAFRVPT NVSV+ LT LE P ++I +K+A+E P LK	
Sbjct:	61	AAKAVGLVLPENGLTGMAFRVPTPNVSVVDLTVELEKPTVEEINAALKEAEGPALK	120
NOV88:	273	GILDYTEHHVVSSEFSDTHSSTFNDGAGIALNDHFVKLISCY	315
		GIL YTE +VSS F D HSS F+ A I LND+FVKL++ Y	
Sbjct:	121	GILGYTEDPLVSSDFIGDPHSSIFDAKATIVLNDNFVKLVAWY	163

NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3-phosphate. It is a cytoplasmic protein that is active as a homodimer, with each monomer containing an N-terminal NAD binding site. In insects, it acts in conjunction with a mitochondrial alpha-glycerophosphate oxidase in the alpha-glycerophosphate cycle, which is essential for the production of energy used in insect flight.

Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.2.12) (GAPDH) mRNA levels, protein, and enzymatic activity increase in 3T3-F442A adipocytes after exposure to physiological concentrations of insulin (Alexander, M., Curtis, G., Avruch, J., and Goodman, H. (1985) J. Biol. Chem. 260, 11978-11985). In order to understand the mechanism of this regulation, researchers isolated and sequenced 5.4 kilobase pairs of a 12-kilobase pair human genomic clone encoding a functional GAPDH gene. The gene consists of 9 exons and 8 introns with eukaryotic signals necessary for the transcription and translation of GAPDH mRNA. The exon sequence confirms previously published cDNA sequences for human GAPDH in muscle, liver, and erythrocytes. The organization of the human and the unique chicken GAPDH genes is strikingly similar. Although chicken exons VIII-XI have been fused into human exon 8, introns which separate exons encoding the NAD binding, catalytic, and helical domains of the GAPDH protein have been retained. Stable transfection of rodent cells with the intact human GAPDH gene resulted in the expression of a correctly initiated human GAPDH mRNA and an enzymatically active human GAPDH polypeptide. Thus, the gene contains a functional promoter and intact coding sequences. Although many processed GAPDH pseudogenes and GAPDH-like sequences are present in the human genome, Southern blot analysis of human genomic DNA using a probe derived from the 3'-untranslated region of the GAPDH gene detected only two genes, a 10-copy processed pseudogene and a single copy of the isolated gene. In contrast, a probe derived from an intron segment of the isolated gene detected only a single copy of the GAPDH gene. Collectively, these findings strongly suggest that the human genome encodes a single functional GAPDH gene.

Hopkinson et al. (1974) presented evidence that glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) is a dimer of dissimilar subunits. Electrophoretic variants at each of two loci, designated GPD1 and GPD2, were described. By the method of somatic cell hybridization, Kielty and Povey (1982) assigned the presumed structural gene for alpha-glycerophosphate dehydrogenase to chromosome 12. Since this is a liver-specific enzyme, a rat hepatoma cell line was used as one of the 'parents' in the hybridization.

NOV88 is predicted to be expressed in at least the following tissues: liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV88 is provided in Example 2.

The NOV88 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation as well as other diseases, disorders and conditions. NOV88 nucleic acids encoding the Glycerol-3-Phosphate Dehydrogenase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a Glycerol-3-Phosphate Dehydrogenase-like protein includes the nucleic acid whose sequence is provided in Table 88A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 88A while still encoding a protein that maintains its Glycerol-3-Phosphate Dehydrogenase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 88A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 13% of the bases may be so changed.

The novel protein of the invention includes the Glycerol-3-Phosphate Dehydrogenase-like protein whose sequence is provided in Table 88B. The invention also includes a mutant or

variant protein any of whose residues may be changed from the corresponding residue shown in Table 88B while still encoding a protein that maintains its Glycerol-3-Phosphate Dehydrogenase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 18% of the amino acid residues may be so changed.

5 These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

10

NOV89

The disclosed NOV89 (alternatively referred to herein as CG56396-01) includes the 1221 nucleotide sequence (SEQ ID NO:287) shown in Table 89A. A NOV89 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 25-27 and ends with a stop codon at nucleotides 1033-1035. The disclosed NOV89 maps to human chromosome 6.

15

Table 89A. NOV89 Nucleotide Sequence (SEQ ID NO:287)

```
CTGCATCTTCTCGTGCATTGCCAGCTGCATCCCTGAGACACCATGGTGAAGGTGAAGGCTGGAGTCAACA
GATTTGGTTGTATTGGCTGCTGGTCACCAGGGCTGCTTTAAACTCTGGTTTAGTCGATATTGTCGCCAT
CAATGACCCCTTCATTGACCTCAACAACACTGTCTACATGTTCCAGTATAATTCGCCCATGGCAAATTC
CACGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTATCAATGGAATCTCATCACTATTTTCCAGGGGC
AAGATCTCACCAAAATCAAATGGGGCAATGCTGGCACTGAGTACATCATGGAGTTCACCAGCATCTTCAC
CACCATGGAGAAGGCTGGGGCTCACTTGGAGGGAGGAGCCAAAACGGTCATCATCTGACCCCTGCTGCT
GATGCCCCCATGTTTCGTGATGGGTGTGAACCATGAGAAATATGACAACAGCTCAAGATTACTCAAGATTA
TCAGCAATGCCTCCTGCACCACCAGCTGCTTAACGCCCCCTGGCCAAGGTCATCCATGACAACCTTGGTAC
CGTGGAAGGACTCATGACCATCGCTGCCACCCAGAACTATGGATGGCTCCTATGGGAACTGTGGGGT
GACGGCCATGGGGCTCTCCAGAACATCCTCTGCTCTACTGGTGCTGCCAAGGCTGTGAGGAAGGTCA
TCCCTGAGCTAAACGGGAAGCTCACTGGCATGGCCTTCCGTGTCCTCACTGCCAACATGTAGTGGTGA
CCTGACCTGCCGTCTGAAAAACCTACCAATATGATGACACCAAGAAGGTGGTGAAGCAGGCGTCAGAG
GACCCCTCAAAGGCATCCTGGGCTACTCTGAGCACCAGGTGGTCTCTCCAACTTCAACTCAACAGACA
CCCCTCTTCCACCTTCGATGCTGGGGCTGGCATTGCCCTCAACGACCACTTGTCAAGCTCATTTCCTG
GTATGACAAATGAATTGGCTGCAGCAACAGGGTGGTGGACCTCTGCCACAGTGTGGCTTCAAGGAGTA
AGACCCCGAGACCAGCCCGACGACGACGACGAGGGAAGAGAGCGGCCCTCACTGCTGGAGAGTCCC
TGCCCACTCAGTCTCCACCACTGAGAATCTCCCTCCTCATAGTTTCCATGCAGACCCCTAAAAG
GGAGGAGCCGAGGGAGCCCCACCTTTTCATG
```

A NOV89 polypeptide (SEQ ID NO:288) encoded by SEQ ID NO:287 is 374 amino acids in length and is presented using the one-letter amino acid code in Table 89B. The Psort profile for NOV89 predicts that this sequence is likely to be localized, to the endoplasmic reticulum (membrane) with a certainty of 0.5500, or to lysosomes with a certainty of 0.2630.

20

Table 89B. NOV89 Polypeptide Sequence (SEQ ID NO:288)

```
MVKVKAGVNRFGCIGCLVTRAAALNSGLVDIVAINDPFIDLNNTVYMFQYNSAHGKFHGTV
KAENGKLVINGNLITIFQGQDLTKIKWGNAGTEYIMEFTSIFTTMEKAGAHLEGGAKTVI
```


ISAPSADAPMFVMGVNHEKYDNSRLLKII SNASCTTSCLTPLAKVIHDNFGTVEGLMTI
AATQKTMDSYGKLGWDGHGALQNILSASTGAAKAVRKVIPELNGKLTGMAFRVPTANMS
VVDLTCRLEKPTKYDDTKKVVQASEDPLKGILGYSEHQVSSNFNSTDTHSSTFDAGAG
IALNDHFVKLISWYDNEFGCSNRVVDLCPQCGFGQVRPPDHQPQRQHDGKRAALTAGESL
PHSVSHHTENLPSS

A BLAST analysis of NOV89 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV89 had high homology to other proteins as shown in Table 89C.

5

Table 89C. BLASTX results from PatP database for NOV89			
Sequences producing High-scoring Segment Pairs:		High Score	Smallest Sum Probability P (N)
patp:AA07036	Breast cancer associated antigen precursor	1393	3.0e-142
patp:AA05368	Human HCMV inducible gene protein	1393	3.0e-142
patp:AAG64817	Human G3PDH fragment - <i>Homo sapiens</i> , 327 aa.	1377	1.5e-140
patp:AAE04373	Mouse cancer associated antigen OY-MC-2	1315	5.6e-134
patp:AAR12995	GAP-DH - <i>Aspergillus oryzae</i> (ATCC 42149	984	6.6e-99

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1066 of 1225 bases (87%) identical to a gb:GENBANK-ID:AF261085[acc:AF261085.1 mRNA from *Homo sapiens* (glyceraldehyde-3-phosphate dehydrogenase (GADPH) mRNA). The full amino acid sequence of the protein of the invention was found to have 279 of 327 amino acid residues (85%) identical to, and 293 of 327 amino acid residues (89%) similar to, the 335 amino acid residue ptnr:TREMBLNEW-ACC:AAG01996 protein from *Homo sapiens* (Human) (CLONE CDABP0047 MRNA SEQUENCE). NOV89 also has homology to the other proteins shown in the BLASTP data in Table 89D.

10

15

Table 89D. NOV89 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 7669492 ref NP_002037.2 (NM_002046)	glyceraldehyde-3-phosphate dehydrogenase [<i>Homo sapiens</i>]	335	279/331 (84)	293/331 (88)	e-153
gi 31645 emb CA A25833.1 (X01677)	glyceraldehyde-3-phosphate dehydrogenase [<i>Homo sapiens</i>]	335	278/331 (83)	293/331 (87)	e-152
gi 2407184 gb A AB94053.1 (AF017079)	glyceraldehyde 3-phosphate dehydrogenase [<i>Sus scrofa</i>]	333	274/328 (83)	289/328 (87)	e-149

gi 6983849 dbj BAA90818.1 (AB038241)	glyceraldehyde-3- phosphate dehydrogenase [Felis catus]	333	268/328 (81)	283/328 (85)	e-146
gi 2506441 sp P 00355 G3P_PIG	GLYCERALDEHYDE 3- PHOSPHATE DEHYDROGENASE (GAPDH)	333	267/328 (81)	283/328 (85)	e-146

This BLASTP data is displayed graphically in the ClustalW in Table 89E. A multiple sequence alignment is given, with the NOV89 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 89D.

Table 89E. ClustalW Alignment of NOV89

NOV89	(SEQ ID NO:288)
gi 7669492	(SEQ ID NO:744)
gi 31645	(SEQ ID NO:745)
gi 2407184	(SEQ ID NO:746)
gi 6983849	(SEQ ID NO:747)
gi 2506441	(SEQ ID NO:748)

	10	20	30	40	50
NOV89	MVKV	KAGV	NRFC	IGCL	VTAA
gi 7669492	MVKV	KAGV	NRFC	IGCL	VTAA
gi 31645	MVKV	KAGV	NRFC	IGCL	VTAA
gi 2407184	--MV	KAGV	NRFC	IGCL	VTAA
gi 6983849	--MV	KAGV	NRFC	IGCL	VTAA
gi 2506441	--MV	KAGV	NRFC	IGCL	VTAA

	60	70	80	90	100
NOV89	SAHG	KFHG	TVKA	ENGK	LVI
gi 7669492	SAHG	KFHG	TVKA	ENGK	LVI
gi 31645	SAHG	KFHG	TVKA	ENGK	LVI
gi 2407184	SAHG	KFHG	TVKA	ENGK	LVI
gi 6983849	SAHG	KFHG	TVKA	ENGK	LVI
gi 2506441	SAHG	KFHG	TVKA	ENGK	LVI

	110	120	130	140	150
NOV89	VFTT	MEKA	GAHL	GGAK	RVII
gi 7669492	VFTT	MEKA	GAHL	GGAK	RVII
gi 31645	VFTT	MEKA	GAHL	GGAK	RVII
gi 2407184	VFTT	MEKA	GAHL	GGAK	RVII
gi 6983849	VFTT	MEKA	GAHL	GGAK	RVII
gi 2506441	VFTT	MEKA	GAHL	GGAK	RVII

	160	170	180	190	200
NOV89	SNAS	CTTN	CLAP	LAKV	IHDN
gi 7669492	SNAS	CTTN	CLAP	LAKV	IHDN
gi 31645	SNAS	CTTN	CLAP	LAKV	IHDN
gi 2407184	SNAS	CTTN	CLAP	LAKV	IHDN
gi 6983849	SNAS	CTTN	CLAP	LAKV	IHDN
gi 2506441	SNAS	CTTN	CLAP	LAKV	IHDN

	210	220	230	240	250
NOV89	SNAS	CTTN	CLAP	LAKV	IHDN
gi 7669492	SNAS	CTTN	CLAP	LAKV	IHDN
gi 31645	SNAS	CTTN	CLAP	LAKV	IHDN
gi 2407184	SNAS	CTTN	CLAP	LAKV	IHDN
gi 6983849	SNAS	CTTN	CLAP	LAKV	IHDN
gi 2506441	SNAS	CTTN	CLAP	LAKV	IHDN

NOV89	DGHGALQNILSASTGAAKAVRKVIPELNGKLTGMAFRVPTANMSVVDLTC
gi 7669492	DGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTC
gi 31645	DGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTC
gi 2407184	DGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVDLTC
gi 6983849	DGRGAQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTPNVSVDLTC
gi 2506441	DGRGAQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTPNVSVDLTC
	260 270 280 290 300
NOV89	RLEKPTKYDDTKKVVKQASEDPLKGILGYSEHQVSSNFNSTDTHSSTFD
gi 7669492	RLEKPAKYDDIKKVVKQASEGPLKGILGYTEHQVSSDFNS-DTHSSTFD
gi 31645	RLEKPAKYDDIKKVVKQASEGPLKGILGYTEHQVSSDFNS-DTHSSTFD
gi 2407184	RLEKPAKYDDIKKVVKQASEGPLKGILGYTEHQVSSDFNS-DTHSSTFD
gi 6983849	RLEKPAKYDDIKKVVKQASEGPLKGILGYTEDQVVSDFNS-DTHSSTFD
gi 2506441	RLEKPAKYDDIKKVVKQASEGPLKGILGYTEDQVVSDFNS-DTHSSTFD
	310 320 330 340 350
NOV89	AGAGIALNDHFVKLISWYDNEFGCSNRVVDLCPQCGFGVRRPPDHQPQRQ
gi 7669492	AGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE-----
gi 31645	AGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE-----
gi 2407184	AGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE-----
gi 6983849	AGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE-----
gi 2506441	AGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE-----
	360 370
NOV89	HDGKRAALTAGESLPHSVSHHTENLPSS
gi 7669492	-----
gi 31645	-----
gi 2407184	-----
gi 6983849	-----
gi 2506441	-----

Table 89F lists the domain description from DOMAIN analysis results against NOV89. This indicates that the NOV89 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 89F. Domain Analysis of NOV89				
gnl Pfam pfam02800, gpdh_C, Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain. GAPDH is a tetrameric NAD-binding enzyme involved in glycolysis and glycconeogenesis. C-terminal domain is a mixed alpha/antiparallel beta fold. SEQ ID NO:881				
CD-Length = 163 residues, 100.0% aligned				
Score = 209 bits (531), Expect = 3e-55				
NOV89:	156	TTSCLTPLAKVIHDNFGTVEGLMTI----	AATQKTMDGSYGKLWGDGHGALQNILSASTG	211
Sbjct:	1	TTNCLAPLAKVLNDNFGIEKGLMTTVHAYTADQKLVDGPHHKDLRRGRAAPNIIPTSTG		60
NOV89:	212	AAKAVRKVIPELNGKLTGMAFRVPTANMSVVDLTCRLEKPTKYDDTKKVVKQASEDP-LK		270
Sbjct:	61	AAKAV V+PELNGKLTGMAFRVPT N+SVVDLT LEKP ++ +K+A+E P LK		120
NOV89:	271	GILGYSEHQVSSNFNFNSTDTHSSTFDAGAGIALNDHFVKLISWY		314
Sbjct:	121	GILGYTE+VSS+F D HSS FDA A I LND+FVKL+WY		163

NAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) (GPD) catalyzes the reversible reduction of dihydroxyacetone phosphate to glycerol-3-phosphate. It is a cytoplasmic protein that is active as a homodimer, each monomer containing an N-terminal NAD binding site. In insects, it acts in conjunction with a mitochondrial alpha-glycerophosphate oxidase in the alpha-glycerophosphate cycle, which is essential for the production of energy used in insect flight.

Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.2.12) (GAPDH) mRNA levels, protein, and enzymatic activity increase in 3T3-F442A adipocytes after exposure to physiological concentrations of insulin (Alexander, M., Curtis, G., Avruch, J., and Goodman, H. (1985) *J. Biol. Chem.* 260, 11978-11985). In order to understand the mechanism of this regulation, researchers isolated and sequenced 5.4 kilobase pairs of a 12-kilobase pair human genomic clone encoding a functional GAPDH gene. The gene consists of 9 exons and 8 introns with eukaryotic signals necessary for the transcription and translation of GAPDH mRNA. The exon sequence confirms previously published cDNA sequences for human GAPDH in muscle, liver, and erythrocytes. The organization of the human and the unique chicken GAPDH genes is strikingly similar. Although chicken exons VIII-XI have been fused into human exon 8, introns which separate exons encoding the NAD binding, catalytic, and helical domains of the GAPDH protein have been retained. Stable transfection of rodent cells with the intact human GAPDH gene resulted in the expression of a correctly initiated human GAPDH mRNA and an enzymatically active human GAPDH polypeptide. Thus, the gene contains a functional promoter and intact coding sequences. Although many processed GAPDH pseudogenes and GAPDH-like sequences are present in the human genome, Southern blot analysis of human genomic DNA using a probe derived from the 3'-untranslated region of the GAPDH gene detected only two genes, a 10-copy processed pseudogene and a single copy of the isolated gene. In contrast, a probe derived from an intron segment of the isolated gene detected only a single copy of the GAPDH gene. Collectively, these findings strongly suggest that the human genome encodes a single functional GAPDH gene.

Hopkinson et al. (1974) presented evidence that glycerol-3-phosphate dehydrogenase (EC 1.1.1.8) is a dimer of dissimilar subunits. Electrophoretic variants at each of two loci, designated GPD1 and GPD2, were described. By the method of somatic cell hybridization, Kielty and Povey (1982) assigned the presumed structural gene for alpha-glycerophosphate dehydrogenase to chromosome 12. Since this is a liver-specific enzyme, a rat hepatoma cell line was used as one of the 'parents' in the hybridization.

NOV89 is predicted to be expressed in at least the following tissues: liver. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV89 is provided in Example 2.

5 The NOV89 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, cirrhosis, transplantation as well as other diseases, disorders and conditions. NOV89 nucleic acids encoding the Glycerol-3-Phosphate Dehydrogenase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications,
10 wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a Glycerol-3-Phosphate Dehydrogenase-like protein includes the nucleic acid whose sequence is provided in Table 89A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 89A while still
15 encoding a protein that maintains its Glycerol-3-Phosphate Dehydrogenase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 89A, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements
20 thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant
25 nucleic acids, and their complements, up to about 13% of the bases may be so changed.

The novel protein of the invention includes the Glycerol-3-Phosphate Dehydrogenase-like protein whose sequence is provided in Table 89B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 89B while still encoding a protein that maintains its Glycerol-3-Phosphate
30 Dehydrogenase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 15% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using

prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below.

NOV90

- 5 The disclosed NOV90 (alternatively referred to herein as CG56888-01) includes the 1686 nucleotide sequence (SEQ ID NO:289) shown in Table 90A. A NOV90 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 58-60 and ends with a stop codon at nucleotides 1630-1632. The disclosed NOV90 maps to human chromosome 3.

Table 90A. NOV90 Nucleotide Sequence (SEQ ID NO:289)

GTGCCTCACTGGGGTGGGGAACCTGCCTCACCTGGGGCCATTTTCATATTTCTGAATCATGTGTGATAACA
GAGAAGCTGGAAGACAAGCCTCCAGCACCTCCCGTGCAGATGAGCAGGACCATCTTTAGCACTGGGGGCAA
AGACCTTTTGTCCAGCAATCACAGTTTGAACCTTTGCCTTCTGTTCCAGAGGAGAAAAAGCCAGGCAT
AAAATCATCTCCATATTTCTCAGGCACAGAGAAAGGAAGTAAAAAGAAAGAAAGGAACGCCAGAAATTT
CTCCTCCATCTGATTTTGAACACACCATCCATGTTGGCTTTGATGCTGTTACTGGAGAATTCACCTGGCAT
GCCAGAACAGTGGGCTCGATTACTACAGACCTCCAATATCACCAACTACAGCAAAAGAAGAATCCTCAG
GCTGTGCTGGATGCTACGACTCCAACACAGTGAAGCAGAAGTATCTGAGTTTACTCCTCTGAGAAAG
ATGGCTTCCCTTCTGGAACACCAGCACTGAATGCCGAGGGAACAGAAGCACCTGCAGTAGTGACAGAGGA
GGAGGACGATGATGAAGAGACTGCCCTCCCATTTATTGCCCCACCACCGGATCATATGAAATCAATTTAC
ACACGGTCTGTAATTGACCTGTTCTGTCACCACTGTTGGTGATTCAAATGTTGATGGTGGTCCAAGTCTT
TAGACAAACAGAAAAAGAACTAAGATGACAGATGAAGAGATTATGGAGAACTAAGAACTATTGTGAG
CATAGGTGACCCTAAGAAAAAAGAAAAAATATACAAGATATGAAAAAATTGGACAAGGGGCTTCTGGT
ACAGTTTTTCACTGCTACTGACGTTGCACTGGGACAGAAGGTTGCTATCAAACAAATTAATTTACAGAAAC
AGCCAAAGAAGGAATTGATCATTAAATGAGATTCTGGTAATGAAAGAATTAAAAAATCCCAACATAGTTAA
CTTCTTGGACAGTTACCTGGTAGGAGATGAATTGTTTGGTGGTGGTGAATACCTTGGTGGGGGTCACCTC
ACTGATGTGGTAACAGAAACCTGCATGGATGAAGCACAGATTGCCGCTGTATGCAGAGAGAGTTTACAGG
CAITGGAGTTTTTACATGCTAATCAAGTGATCCACAGAGACATCAAAAGTGACAGTGACTTTTGGGAAT
GGAAGGATCGGTTAAGCTCACTGACTTTGGTTTCTGTGCCCAGATCACCCTGAGCAGAGCAACCGCAGT
ACCGTGGTCAGAACGCCATACTGGATGGCACCAGAAGTGGTTACACGGAAGGCTTATGGCCCTAAAGTCA
ATGTATGGTCTCTGGGTATCATGGCTACTGAGATGGTAGAAGGAGAGCCTCCATACCTCAATGAAAATCC
CTTGAGGGCCTTGTGCCTAATAGCACTAATGGAATCCAGAACTTCAGAATCCAGAGACACTTTCCCA
ATATTTCGGGATTTCTTAAATCGATGTTTGGAAACAGATGTGGAAGAAAGGGGTTAGCCAAAGAATTAT
TACAGCATCTTTCTGAACTAGCCAACTGTTATCTAGCTTGACACCACTGATCATGGCAGCTAAAGA
AGCAATGAAGAGTAACCGTTAATCACTGCTGTGGCCTCATATTCTTTTCCATTTCTACAAGAAGC
CTTTTA

- 10 A NOV90 polypeptide (SEQ ID NO:290) encoded by SEQ ID NO:289 is 524 amino acids in length and is presented using the one-letter amino acid code in Table 90B. The Psort profile for NOV90 predicts that this sequence is likely to be localized to the nucleus with a certainty of 0.7000.

15

Table 90B. NOV90 Polypeptide Sequence (SEQ ID NO:290)

MCDNRELEDKPPAPPVVRMSRTIFSTGGKDP LSAHSLKPLPSVP EKKPRHKIISIFSGT
EKSGKKKEKERPEISPPSDFEHTIHVGFDV TGEFTGMPEQWARLLQTSNITKLQOKKNP
QAVLDVYDSNTVKQKYLSTPPEKDFPSPGTPALNAEGTEAPAVVTEEDDDDEETAPPII
APPPDHMKSIYTRSVIDPVPAPVGDSDNDGGAKSLDKQKKTKMTDEEIMEKLRTIVSIG
DPKKRKKYTRYEKIGQGASGTVFTATDVALGQKVAIKQINLQKQPKKELIINEILVMKE
LKNPNIVNFDLSYLVGDELFFVVEYLARGSLTDVVTETCMDEAQIAAVCRESLQALEFLH
ANQVIHRDIKSDSVLLGMEGSVKLTDGFCQAITPEQSKRSTVVRTPYWMAPEVVTRKAY
GPKVNVWSLGMATEMVEGEPYLNENPLRALCL IATNGIPELQNPETLSPIFRDFLNRCL
LETDVEKRGSAKELLQHLFLKLAKLLSSLTPLIMAAKEAMKSNR

A BLAST analysis of NOV90 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV90 had high homology to other proteins as shown in Table 90C.

5

Table 90C. BLASTX results from PatP database for NOV90			
Sequences producing High-scoring Segment Pairs:		High Score	Smallest Sum
			Probability P(N)
patp:AAB03969	p-21 activated protein kinase (PAK2)	2494	6.5e-259
patp:AAW95521	Protease activated protein kinase I (PAK I)	2459	3.3e-255
patp:AAR96296	Human p21-protein activated serine kinase	2361	8.0e-245
patp:AAW13379	Human p21 activated serine kinase p65 protein	2361	8.0e-245
patp:AAW47119	Human p21-activated serine kinase p65	2361	8.0e-245

In a search of public sequence databases, it was found, for example, that the amino acid sequence of the NOV90 protein of the present invention was found to have 503 of 524 amino acid residues (95%) similar to, the 524 amino acid residue ptr:SWISSNEW-ACC:Q13177 protein from *Homo sapiens* (Human) (SERINE/THREONINE-PROTEIN KINASE PAK 2 (EC 2.7.1.-) (P21-ACTIVATED KINASE 2) (PAK-2) (PAK65) (GAMMA-PAK) (S6/H4 KINASE)). NOV90 also has homology to the other proteins shown in the BLASTP data in Table 90D.

10

Table 90D. NOV90 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 3041712 sp Q13177 PAK2_HUMAN	SERINE/THREONINE-PROTEIN KINASE PAK 2 (P21-ACTIVATED KINASE 2) (PAK-2) (PAK65) (GAMMA-PAK) (S6/H4 KINASE)	524	493/527 (93)	503/527 (94)	0.0
gi 4505599 ref NP_002568.1 (NM_002577)	p21 (CDKN1A)-activated kinase 2; novel serine kinase; hPAK65 [<i>Homo sapiens</i>]	525	493/528 (93)	503/528 (94)	0.0
gi 16758002 ref NP_445758.1 (NM_053306)	p21 (CDKN1A)-activated kinase 2 [<i>Rattus norvegicus</i>]	524	483/527 (91)	498/527 (93)	0.0
gi 2499647 sp Q29502 PAK2_RABBIT	SERINE/THREONINE-PROTEIN KINASE PAK 2 (P21-ACTIVATED KINASE 2) (PAK-2) (GAMMA-PAK) (P21-ACTIVATED PROTEIN KINASE I) (PAKI)	524	484/527 (91)	498/527 (93)	0.0
gi 6288680 gb AAF06695.1 U19967.1 (U19967)	PAK2 [<i>Rattus norvegicus</i>]	524	481/527 (91)	498/527 (94)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 90E. A multiple sequence alignment is given, with the NOV90 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 90D.

Table 90E. ClustalW Alignment of NOV90	
NOV90	(SEQ ID NO:290)
gi 3041712	(SEQ ID NO:749)
gi 4505599	(SEQ ID NO:750)
gi 16758002	(SEQ ID NO:751)
gi 2499647	(SEQ ID NO:752)
gi 6288680	(SEQ ID NO:753)
	10 20 30 40 50
NOV90	MCDNRELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
gi 3041712	MSDNGELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
gi 4505599	MSDNGELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
gi 16758002	MSDNGELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
gi 2499647	MSDNGELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
gi 6288680	MSDNGELEDKPPAPPVVRMSSTIFSTGGKDPLSANHSLKPLPSVPEKKPR
	60 70 80 90 100
NOV90	HKIISIFSGTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
gi 3041712	HKIISIFSGTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
gi 4505599	HKIISIFSGTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
gi 16758002	NKIISIFSSTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
gi 2499647	NKIISIFSGTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
gi 6288680	NKIISIFSSTEKSGSKKKEKERPEISPPSDFEHTIHVGFDVAVTGEFTGMPE
	110 120 130 140 150
NOV90	QWARLLQTSNITKLEQKKNPQAVLDV--YDSNTVKQKYLSTPPEKDG
gi 3041712	QWARLLQTSNITKLEQKKNPQAVLDVLFYDSNTVKQKYLSTPPEKDG
gi 4505599	QWARLLQTSNITKLEQKKNPQAVLDVLFYDSNTVKQKYLSTPPEKDG
gi 16758002	QWARLLQTSNITKLEQKKNPQAVLDVLFYDSNTVKQKYLSTPPEKDG
gi 2499647	QWARLLQTSNITKLEQKKNPQAVLDVLFYDSNTVKQKYLSTPPEKDG
gi 6288680	QWARLLQTSNITKLEQKKNPQAVLDVLFYDSNTVKQKYLSTPPEKDG
	160 170 180 190 200
NOV90	PSGTPALNAEGTEAPAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
gi 3041712	PSGTPALNAKGTEAPAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
gi 4505599	PSGTPALNAKGTEAPAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
gi 16758002	PSGTPALNTKGSETSAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
gi 2499647	PSGAPALNTKGSETSAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
gi 6288680	PSGTPALNTKGSETSAVVTEEDDDDEETAPPVIAAPPDHMKSIYTRSV
	210 220 230 240 250
NOV90	PVPAPVGDSNVDSGAKSLDKQKKKMTDEEIMEKLRTIVSIGDPKKK
gi 3041712	PVPAPVGDSHVDGAKSLDKQKKKMTDEEIMEKLRTIVSIGDPKKK
gi 4505599	PVPAPVGDSHVDGAKSLDKQKKKMTDEEIMEKLRTIVSIGDPKKK
gi 16758002	PVPAPVGDSNVDSGAKSSDKQKKKAKMTDEEIMEKLRTIVSIGDPKKK
gi 2499647	PVPAPVGDSHVDGAKSSDKQKKKMTDEEIMEKLRTIVSIGDPKKK
gi 6288680	PVPAPVGDSNVDSGAKSSDKQKKKAKMTDEEIMEKLRTIVSIGDPKKK
	260 270 280 290 300

Score = 251 bits (640), Expect = 1e-67					
NOV90:	255	IGQGASGTVFTATDVALGQKVAIKQINLQKQPKK--ELIINEILVMKELKNPNIVNFLDS	312		
		+G+GA G V+ A D G+ VAIK I +K KK E I+ EI ++K+L +PNIV D			
Sbjct:	7	LGKGAFGKVYLARDKKTGKLVAIKVIKKEKLKKKKRERILREIKILKKLDHPNIVKLYDV	66		
NOV90:	313	YLVGDELFFVVEYLRAGSLTDVVTET-CMDEAQIAAVCRESLQALEFLHANQVHRDIKS	371		
		+ D+L++V+EY G L D++ + + E + R+ L ALE+LH+ +IHRD+K			
Sbjct:	67	FEDDDKLYLVMEYCEGGDLFDLLKKRGRlseDEARFYARQILSALEYLHSQGI IHRDLKP	126		
NOV90:	372	DSVLLGMEGSVKLTDFGFCQITPEQSKRSTVVRTPYWMAPEVVTRKAYGPKVNVWSLGI	431		
		+++LL +G VKL DFG Q+ + +T V TP +MAPEV+ K YG V++WSLG+			
Sbjct:	127	ENILLDSGDGHVKLADFLAKQLDSGGTLLTTFVGTPEYMAPEVLLGKGYGKAVDIWSLGV	186		
NOV90:	432	MATEMVEGEPFYLNNENPLRALCLLATNGIPELQNPETLSPIFRDFLNRCLETDVKEGRGS	490		
		+ E++ G+PP+ ++ L AL P PE +SP +D + + L D EKR +			
Sbjct:	187	ILYELLTGKPPFPFGDQLLALFKICKPPPPPPPEWKISPEAKDLIKLLVKDPEKRLT	246		
NOV90:	491	AKE	493		
		A+E			
Sbjct:	247	AEE	249		

Serine/threonine kinases are an extensive family of enzymes that catalyzes the phosphorylation of serine or threonine residues on its target protein. Protein kinases share a conserved catalytic core common to both serine/ threonine and tyrosine protein kinases. This domain contains residues, which are specific to the distinct types of protein kinases

The S6/H4 kinase purified from human placenta catalyzes phosphorylation of the S6 ribosomal protein, histone H4, and myelin basic protein. In vitro activation of the p60 S6/H4 kinase requires removal of an autoinhibitory domain by mild trypsin digestion and autophosphorylation of the catalytic domain (p40 S6/H4 kinase). The two autophosphorylation/autoactivation sites contain the sequences SSMVGTPY (site 1) and SVIDPVPAPVGDSHVDGAAK (site 2). These sequences identify S6H4 kinase as the rac-activated PAK65 (Martin, G. A., Bollag, G., McCormick, F. and Abo, A. (1995) EMBO J. 14, 1971-1978). Site 1 phosphorylation is most rapid, but activation does not occur until site 2 is autophosphorylated. The site 1 phosphorylation occurs by an intramolecular mechanism whereas site 2 autophosphorylation occurs by an intermolecular mechanism. A model is proposed in which phosphorylation of sites 1 and 2 occurs sequentially. The model proposes that trypsin treatment of the inactive holoenzyme removes an inhibitory rac-binding domain which blocks MgATP access to the catalytic site. The pseudosubstrate domain at site 1 is autophosphorylated and subsequent bimolecular autophosphorylation at site 2 fully opens the catalytic site. Phosphorylation by a regulatory protein kinase may occur at site 2 in vivo.

NOV90 is predicted to be expressed in at least the following tissues: brain, cerebellum, skeletal muscle, ovary, thymus and spleen. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to

SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV90 is provided in Example 2.

The NOV90 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von
5 Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation,
10 immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV90 nucleic acids encoding the PAK2-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a serine/threonine-protein kinase
15 PAK 2-like protein includes the nucleic acid whose sequence is provided in Table 90A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 90A while still encoding a protein that maintains its serine/threonine-protein kinase PAK 2-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes
20 nucleic acids whose sequences are complementary to the sequence of Table 90A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar
25 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

The novel protein of the invention includes the serine/threonine-protein kinase PAK 2-like protein whose sequence is provided in Table 90B. The invention also includes a mutant or
30 variant protein any of whose residues may be changed from the corresponding residue shown in Table 90B while still encoding a protein that maintains its serine/threonine-protein kinase PAK 2-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 6% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV91

The disclosed NOV91 (alternatively referred to herein as CG56779-01) includes the 404 nucleotide sequence (SEQ ID NO:291) shown in Table 91A. A NOV91 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 25-27 and ends with a stop codon at nucleotides 373-375. The disclosed NOV91n maps to human chromosome 3.

Table 91A. NOV91 Nucleotide Sequence (SEQ ID NO:291)

AATGGGTCTCTTTTCTACAGCCCATGCCCTTCCTAGAGCTAGACACGAACCTGCCTGCCAACCAAGTGCTGTCAGGGTTGGAGAAATGGCTCTGCGCCACAGCCTCCATCCTGGGCAAACCAAGGATCATGTGAACATGATGGGTGTAGCGGGCTGACCATGGTGCTGAGTAGGTCCACTGAGCCCTGGGCGCAGCTGTTTCATCTCTCCACCAGCATGATGGACACCACTGAGGAGAACCGCAGCCACAGCAACCACTCTTCGAGTTCCTCACCGAGGAGCTGGCCCTGGGCCAGGACCAGATAATTTTCCACTTTTCCCCCTGGAGCCCTGGCAGACTGGCAA GAAGGGGATGGTCATAACTTTTTAGTGACTGGCCTCGAGGGATCCAGGGCATCT

A NOV91 polypeptide (SEQ ID NO:292) encoded by SEQ ID NO:291 is 116 amino acids in length and is presented using the one-letter amino acid code in Table 91B. The Psort profile for NOV91 predicts that this sequence has no signal peptide and is likely to be localized to microbodies with a certainty of 0.6400. In alternative embodiments, a NOV91 polypeptide is located to the cytoplasm with a certainty of 0.4500.

Table 91B. NOV91 Polypeptide Sequence (SEQ ID NO:292)

MPFLELDTNLPANQVPAGLEKWLKATASILGKPKDHNMMGVAGLTMVLSRSTEPWAQLFISSTSMMDTTEENRSHSTHFFFLTEELALGQDQIIFHFSPLPWPQTGKKGMVITF

A BLAST analysis of NOV91 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV91 had high homology to other proteins as shown in Table 91C.

Table 91C. BLASTX results from PatP database for NOV91

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probabability P(N)

patp:AAR83048	Human macrophage migration inhibitory factor	399	6.5e-37
patp:AAV44997	Human D-dopachrome tautomerase (DDT)	369	9.8e-34
patp:AAB43733	Human cancer associated protein sequence	262	2.1e-22
patp:AAM22110	Peptide #8544 encoded by probe	183	5.0e-14
patp:AAM38563	Peptide #12600 encoded by probe	183	5.0e-14

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 283 of 289 bases (97%) identical to a gb:GENBANK-ID:AP000500|acc:AP000500.1 mRNA from *Homo sapiens* (genomic DNA, chromosome 3p21.3, clone:603 to 320, anti-oncogene region, section 3/3). The full amino acid sequence of the protein of the invention was found to have 82 of 117 amino acid residues (70%) identical to, and 91 of 117 amino acid residues (77%) similar to, the 118 amino acid residue ptmr:pir-id:JE0162 protein from human (dopachrome Delta-isomerase (EC 5.3.3.12)). NOV91 also has homology to the other proteins shown in the BLASTP data in Table 91D.

10

Table 91D. NOV91 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 4503291 ref NP_001346.1 (NM_001355)	D-dopachrome tautomerase [<i>Homo sapiens</i>]	118	82/117 (70)	91/117 (77)	1e-38
gi 4699610 pdb 1DPT A	Chain A, D-Dopachrome Tautomerase	117	81/116 (69)	90/116 (76)	5e-38
gi 7512375 pir G02438	D-dopachrome tautomerase - human	118	80/117 (68)	89/117 (75)	7e-38
gi 13162287 ref NP_077045.1 (NM_024131)	D-dopachrome tautomerase [<i>Rattus norvegicus</i>]	118	71/117 (60)	94/117 (79)	6e-34
gi 6753618 ref NP_034157.1 (NM_010027)	D-dopachrome tautomerase [<i>Mus musculus</i>]	118	67/117 (57)	92/117 (78)	2e-32

This BLASTP data is displayed graphically in the ClustalW in Table 91E. A multiple sequence alignment is given, with the NOV91 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 91D.

15

Table 91E. ClustalW Alignment of NOV91										
NOV91	(SEQ ID NO:292)									
gi 4503291	(SEQ ID NO:754)									
gi 4699610	(SEQ ID NO:755)									
gi 7512375	(SEQ ID NO:756)									
gi 13162287	(SEQ ID NO:757)									
gi 6753618	(SEQ ID NO:758)									
		10	20	30	40	50				
					

NOV91	MPFLELDTNLPANQVPAGLEKWLCA-TASILGKPKDHVNMVGAGLTMVL
gi 4503291	MPFLELDTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMAL
gi 4699610	-PFLELDTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMAL
gi 7512375	MPFLELDTNLPANRVPAGLEKRLCAAAASILGKPADRVNVTVRPGLAMAL
gi 13162287	MPFVELETNLPASRIIPAGLENRLCAATATILDKPEDRVSVTIRPGMTILM
gi 6753618	MPFVELETNLPASRIIPAGLENRLCAATATILDKPEDRVSVTIRPGMTILM
	60 70 80 90 100
NOV91	SRSTEPWAQLFISSTSMMDTTEENRSHSTHFFFEFLTEELALGQDQIIIFHF
gi 4503291	SGSTEPCAQLSISSIGVVGTAEDNRSHSAHFFFEFLTKELALGQDRILIRF
gi 4699610	SGSTEPCAQLSISSIGVVGTAEDNRSHSAHFFFEFLTKELALGQDRILIRF
gi 7512375	SGSTEPCAQCSISSIGVVGTAEDNRSHSAHFFFEFLTKELALGQDRILIRF
gi 13162287	NKSTEPCAHLIISIGVVGTAEQNRSHSSSFFKFLTEELSLDQDRILIRF
gi 6753618	NKSTEPCAHLIISIGVVGTAEQNRSHSASFKEFLTEELSLDQDRIVIRF
	110
NOV91	SPLEPWQTGKKGMVITF-
gi 4503291	FPLESWQIGKIGTVMTFE
gi 4699610	FPLESWQIGKIGTVMTFE
gi 7512375	FPLESWQIGKIGTVMTFE
gi 13162287	FPLEPWQIGKIGTVMTFE
gi 6753618	FPLEAWQIGKIGTVMTFE

Table 91F lists the domain description from DOMAIN analysis results against NOV91. This indicates that the NOV91 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 91F. Domain Analysis of NOV91				
gnl Pfam pfam01187, MIF, Macrophage migration inhibitory factor (MIF). SEQ ID NO:883				
CD-Length = 114 residues, 100.0% aligned				
Score = 126 bits (316), Expect = 8e-31				
NOV91:	2	PFLELDTNLPANQVPAGLEKWLCA-TASILGKPKDHVNMVGAGLTMVLSRSTEPWAQLF	60	
Sbjct:	1	P +DTNLPAN VPAG EK L A A LGKP+D + + G MV ST+P A		
NOV91:	61	ISSTSMMDTTEENRSHSTHFFFEFLTEELALGQDQIIIFHFSPLEPWQTGKKGMVIT	115	
Sbjct:	61	I S ++ E+NRSHS F+FL +EL L +D++ F LE Q G +		
NOV91:	61	IKSIGVV-GAEQNRSHSALLFKFLAKELGLPKDRVYIRFFDLEAAQVGFGNGTTMA	114	

D-Dopachrome tautomerase (DDT) shares a low homologous amino acid sequence (33% homology) with the macrophage migration inhibitory factor (MIF) yet possesses similar tautomerase activity. MIF is a cytokine involved in inflammatory reactions and immune responses. While MIF is a secreted protein, it is not processed from a larger precursor. Whereas recent studies have identified MIF as a pituitary hormone and immunoregulator, less is known about the structural basis of these physiological functions and the real significance of tautomerase activity. D-dopachrome tautomerase, which is related to MIF, is a mammalian cytoplasmic enzyme involved in melanin biosynthesis that tautomerizes 2-carboxy-2,3-

dihydroindole-5, 6-quinone (D-dopachrome) with concomitant decarboxylation to give 5,6-dihydroxyindole (DHI). It is a protein of 117 residues, and acts as a homotrimer.

NOV91 is predicted to be expressed in at least the following tissues: largely in the liver, and to lesser extent in other organs, including the heart, lung, pancreas; and placenta.

5 This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV91 is provided in Example 2.

10 The NOV91 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic
15 thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV91 nucleic acids encoding the D-Dopachrome tautomerase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

20 The novel nucleic acid of the invention encoding a D-dopachrome tautomerase-like protein includes the nucleic acid whose sequence is provided in Table 91A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 91A while still encoding a protein that maintains its D-dopachrome tautomerase-like activities and physiological functions, or a
25 fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to those just described, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications
30 include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In the mutant or variant nucleic acids, and their complements, up to about 3% of the residues may be so changed.

The novel protein of the invention includes the D-dopachrome tautomerase-like protein whose sequence is provided in Table 91B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 91B while still encoding a protein that maintains its D-dopachrome tautomerase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 30% of the bases may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV92

The disclosed NOV92 (alternatively referred to herein as CG56904-01) includes the 1311 nucleotide sequence (SEQ ID NO:293) shown in Table 92A. A NOV92 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 19-21 and ends with a stop codon at nucleotides 1282-1284.

Table 92A. NOV92 Nucleotide Sequence (SEQ ID NO:293)

```
GGAGCTCCACACTTTCAATGGGGAGGCCACCCAGTGGCCGAGCCTGCTGCTGCTCCTGCTGTTGCCGG
GGCCCCCGCCCGTTCGCCGCTTGGAAGACGCTGCCTTCCCCACCTGGGGGAGAGCTTGACGCCCTGCC
CCGGGCTGTCCCTGCGCTGCTCCTGCCCGAGTCGACACTGTGGACTGTGATGGCTTGGACCTTCGA
GTGTTCCCGGACAACATCACCAGAGCCGCTCAGCACCTCTCCCTGCAGAACCAACAGCTCCAGGAACCTC
CCTACAATGAGCTGTCCCGCTCAGTGGCTGCGAACCTCAACCTCCACAACAACCTCATCTCCTCCGA
AGGCCTGCCTGACGAGGCTTCGAGTCCCTACCCAGCTGCAGCACCTCTGCGTGGCTCACAACAAGAAC
AATCTCATCTCCAAGGTGCCCGAGGAGCCCTGAGCCGCCAGACTCAACTCCGTGAGCTCTACCTCCAGC
ACAACCAGCTGACAGACAGTGGCTGGATGCCACCACCTTCAGCAAGCTGCATAGCCTTGAATACCTGGA
TCTCTCCCAACAACCAGCTGACCACAGTGCCCGCGGCTGCCCGGACCTGGCTATCCTGCACCTGGGC
CGCAACCGCATCCGGCAGGTGGAGGCGGCTCGGCTGCACGGGGCGCGTGGTCTGCGCTATTGTTGCTGC
AGCACAACCAGCTGGGAGCTCAGGGCTGCCCGCGGGGCTCTGCGGCCGCTGCGGGGCTGCACACGCT
GCACCTCGATGGCAATGGGCTGGACCGCGTGCCTCCAGCCCTGCCCGCGCGCTGCGTGCCTGGTGTG
CCCCACAACCAGTGGCCGCGCTGGGTGCCCGTGACCTGGTCGCCACACCGGGCCTGACGGAGCTTAAC
TGGCCTATAACCGCTGGCCAGCGCCGCTGTGCACACCGGGCCTTCCGCGGTTGCGTGCCCTGCGCAG
CCTCGACCTGGCAGGAATCAGCTAACCGGCTGCCATGGGCTGCCACTGGCCTGCGCACCTGCAG
CTGCAACGCAACCAGCTGCGGATGTCTGAGCCCGAGCCTCTGGCCGCGCTGGACCAACTGCGGGAGCTCA
GGCTGGCGCACAAACCGGCTCCGGGTGCGGACATCGGGCCAGGCACCTGGCATGAGCTCCAAGCCCTCCA
GGTCAGGCACAGGCTGGTTAGCCACACTGTCCCAGGGCCCTCCATCCCCCTGCCCTGCCACGTC
CCAACATTCTAGTTAGCTGGTAAAGCAATCAGAACAAGAAAATGATAAGA
```

A NOV92 polypeptide (SEQ ID NO:294) encoded by SEQ ID NO:293 is 421 amino acids in length and is presented using the one-letter amino acid code in Table 92B. The Psort profile for NOV92 predicts that this sequence has a signal peptide and is likely to be secreted

with a certainty of 0.4419. The Signal P predicts a likely cleavage site for a NOV92 peptide is between positions 24 and 25, *i.e.*, at the dash in the sequence VAG-LE.

Table 92B. NOV92 Polypeptide Sequence (SEQ ID NO:294)

MGRPTQWPSLLLLLLPGPPPVAGLEDAAFPHLGESLQPLPRACPLRCSCPRVDTVDCDGL
DLRVFPDNITRAAQHLSLQNNQLQELPYNELSRLSGLRTLNLHNNLISSEGLPDEAFESLT
QLQHLCVAHNKNNLISKVPRGALSRTQLRELYLQHNQLTDSGLDATTFSKLHSLEYLDLS
HNQLTTVPAGLPRTLAILHLGRNRIQVEAARLHGARGRLRYLLQHNQLGSSGLPAGALRP
LRGLHTLHLDGNGLDREVPPALPRLRALVLPNHVAALGARDLVATPGLTELNLAYNRLAS
ARVHHRAFRRLRALRSLDLAGNQLTRLPMGLPTGLRTLQLRNQLRMLEPEPLAGLDQLRE
LSLAHNRLRVGDIGPGTWHELQALQVRHRLVSHTVPRAPPSPCLPCHVPNILVSW

- 5 A BLAST analysis of NOV92 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV92 had high homology to other proteins as shown in Table 92C.

Table 92C. BLASTX results from PatP database for NOV92

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)
patp:AAU13396 Amino acid sequence of protein PRO332	690	4.9e-70
patp:AAB33425 Human PRO332 protein	690	4.9e-70
patp:AAB80264 Human PRO332 protein - <i>Homo sapiens</i> , 642 aa.	690	4.9e-70
patp:AAU12356 Human PRO332 polypeptide sequence	690	4.9e-70
patp:AAM41258 Human polypeptide SEQ ID NO 6189	334	5.0e-30

- 10 In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1310 of 1312 bases (99%) identical to a gb:GENBANK-ID:AK027100|acc:AK027100.1 mRNA from *Homo sapiens* (cDNA: FLJ23447 fis, clone HSI03346). The full amino acid sequence of the protein of the invention was found to have 290 of 291 amino acid residues (99%) identical to, and 290 of 291 amino acid residues (99%) similar to, the 363 amino acid residue ptrn:TREMBLNEW-ACC:BAB15657 protein from *Homo sapiens* (Human) (CDNA: FLJ23447 FIS, CLONE HSI03346). NOV92 also has homology to the other proteins shown in the BLASTP data in Table 92D.
- 15

Table 92D. NOV92 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
-------------------------	--------------------	-------------	--------------	--------------	--------

gi 13376224 ref NP_079101.1 (NM_024825)	hypothetical protein FLJ23447 [<i>Homo sapiens</i>]	363	290/291 (99)	290/291 (99)	e-136
gi 11761721 gb AAG40157.1 AF247822.1 (AF247822)	decorin [<i>Oreochromis niloticus</i>]	359	87/281 (30)	137/281 (47)	4e-27
gi 6759315 dbj BAA90246.1 (AB037269)	biglycan [<i>Xenopus laevis</i>]	368	90/289 (31)	134/289 (46)	1e-26
>gi 8134605 sp P21793 PGS2_PIG.	BONE PROTEOGLYCAN II PRECURSOR (PG-S2) (DECORIN)	360	94/298 (31)	144/298 (47)	4e-25
gi 129949 sp P21793 PGS2_B_OVIN	BONE PROTEOGLYCAN II PRECURSOR (PG-S2)	360	88/292 (30)	140/292 (47)	6e-25

This BLASTP data is displayed graphically in the ClustalW in Table 92E. A multiple sequence alignment is given, with the NOV92 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 92D.

Table 92E. ClustalW Alignment of NOV92

NOV92	(SEQ ID NO:294)
gi 13376224	(SEQ ID NO:759)
gi 11761721	(SEQ ID NO:760)
gi 6759315	(SEQ ID NO:761)
gi 8134605	(SEQ ID NO:762)
gi 129949	(SEQ ID NO:763)
NOV9210.....20.....30.....40.....50
NOV92	MGRPTQWPSLLLLLLPGLPPVAGLEDAAPHLGELSLOPLPRACPLRCS
gi 13376224	MAWTFECS-----RTTS-PEELS---TS---PCRTTSSRNSP---T-MSC
gi 11761721	MRSACLSLL---TVTACWALPFR--QSGFLDFMMED-----
gi 6759315	MKVLILLCS---CILVIHALPFE--QRCFWDFSMDDGMAMM-----
gi 8134605	MKATIVFLL---IAQVSWAGPFO--QKGLDFDFMLED-----
gi 129949	MKATIIIFLL---VAQVSWAGPFO--QKGLDFDFMLED-----
NOV9260.....70.....80.....90.....100
NOV92	PRVDTVDCGLDLRVFEDNITRAAQHLISLQNNQLQELFYNELSRLSGLRT
gi 13376224	P--ASVACE-----PS--TSTTTSS--P-----PKACLTRPS---
gi 11761721	---EGSGD-----PVTESPLPPVIG---GPKCPFRQCQCHLRVIOC
gi 6759315	KDEEASGVG-----PIPTESIPDVGLPP---MDLCPFGCQCHLRVVQC
gi 8134605	---EASGIG-----PEDRFPEVPELEPL---GPMCPFRQCQCHLRVVQC
gi 129949	---EASGIG-----PEEHFPEVPELEPM---GPVCPFRQCQCHLRVVQC
NOV92110.....120.....130.....140.....150
NOV92	LNTHNNLTSSEGLFDEAFESLTQLQHLCAHNKNNLISKVPRGALSRTQ
gi 13376224	-----SPS---P-----S---CSTSAWLNHNKNNLISKVPRGALSRTQ
gi 11761721	SDLGLKAVP-EDIP-----DDTLLDLQNNKITEIKENDFKNLKG
gi 6759315	SDLGLTSIP-KNLP-----KDTLLDLQNNKITEIKDDFKGLTN
gi 8134605	SDLGLDKVP-KDLP-----PDTALLDLQNNKITEIKDGFKNLKN
gi 129949	SDLGLEKVP-KDLP-----PDTALLDLQNNKITEIKDGFKNLKN
NOV92160.....170.....180.....190.....200

[illegible]

Leucine-rich repeats (LRRs) are relatively short motifs (22-28 residues in length) found in a variety of cytoplasmic, membrane and extracellular proteins. Although these proteins are associated with widely different functions, a common property involves protein-protein interaction. Little is known about the 3D structure of LRRs, although it is believed that they can form amphipathic structures with hydrophobic surfaces capable of interacting with membranes. In vitro studies of a synthetic LRR from *Drosophila* Toll protein have indicated that the peptides form gels by adopting beta-sheet structures that form extended filaments. These results are consistent with the idea that LRRs mediate protein-protein interactions and

cellular adhesion. Other functions of LRR-containing proteins include, for example, binding to enzymes and vascular repair. The 3-D structure of ribonuclease inhibitor, a protein containing 15 LRRs, has been determined, revealing LRRs to be a new class of alpha/beta fold. LRRs form elongated non-globular structures and are often flanked by cysteine rich domains.

5 NOV92 is predicted to be expressed in at least the following tissues: colon, brain. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV92 is provided in Example 2.

10 The NOV92 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic
15 thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV92 nucleic acids encoding the LRR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

20 The novel nucleic acid of the invention encoding a secreted leucine-rich repeat (LRR) protein-like protein includes the nucleic acid whose sequence is provided in Table 92A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 92A while still encoding a protein that maintains its secreted leucine-rich repeat (LRR) protein-like activities and
25 physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 92A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications
30 include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the secreted leucine-rich repeat (LRR) protein-like protein whose sequence is provided in Table 92B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 92B while still encoding a protein that maintains its Secreted leucine-rich repeat (LRR) protein-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 1% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV93

The disclosed NOV93 (alternatively referred to herein as CG56277-01) includes the 1518 nucleotide sequence (SEQ ID NO:295) shown in Table 93A. A NOV93 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 2-4 and ends with a stop codon at nucleotides 1556-1558. The disclosed NOV93 maps to human chromosome 2.

Table 93A. NOV93 Nucleotide Sequence (SEQ ID NO:295)
CATGAGGGGACTACCTATCAGCAGCAGCACCGGTACGTGGTCGAGGACGGGTTCCTGCGACCGTGCAG CAGCTCTTCGCCAGTGAACAGGGACTACCTACAACGACTTCTTGATTCTCCAGGATTCATAGACTTCA TAGCTGATGAGGTGGACCTGACCTCAGCCCTGACCCAACCGGTCACTCTGAAGACGCCGCTGATCTCCTC CCCCATGGACACTGTGACAGAGGCCGACCTGGCCATCGTGATGGCTCTGATGGGAGGTACTGGTTTCATT CACCACAAGTGCACCCAGAGTTCCAGGCCAGTGAGGTGCAGAAGGTCAAGAAGTTGAACCGGGCTTTA TCACACACCCCGTGGTGCTGAGCCCCCTTGCACACTGTGGGTGATGTGTTGGAGGCCAAGATGCGTCATGG CTTCTCTGGCATCCCCATCACTGAGACGGGTACCATGGGCAGCAAGCTGGTGGGCATCGTCACCTCCCCGA GACATCGACTTTCTTGCTGAGAAGGACCACACCACCTCCTCAGTGAGGTGATGATGCCAAGGATCAAGC TAGTGGTGGCTCCAGCAAGCAGTGTGAGGTTGAAAGAGGGCAAATGAGATCCTGCAGCTTAGTAAGAAAGG AAAGCTGCCTATCGTCAATGATCGCGATGAGCTGGTGGCCATTATCACCTGCACCGCGCTGAAGAACCGA GACTACCCCTGTGGCCTCCAAGGATTCCCATGAGCAGCTGCTGGGCGGGGCAGCTGTGGGTACCCATGAGG ATGACAAATACCACCTGGACCTGCTCACCCAGGTAGGCGTCAATGTCATAGGCTTGGACTCGTCCCAAGG GAACTCGGTGTATCAGATCGCCATGGTGCATTACATCAACAAAAGTACCCCCACCTCCAGGTGATTGGG GGGAACGTGGTGACAGCAGCCCAGGCCAACAACCTGATTGACGCTGGTGTGGATGGGCTGGGCAGGGGCA TGGACTGCGCGGCTGGCTCCATCTACATCAACCAGGAAGTGATAGCCTGCAGTCAGCCCCAGGGCACTGC TGTGTACAAGGTGGCCAAGCATAACCCAGAAGTTTGGTGTGCCATCATAGCCGATGGTGGCATCCAGACC ATGGGGCATGTGTCAAGGCCCTGGCCCTAGGAGCCTCCACAGTGATGATGGGCTCCCTGCTGGCCGCCA CCATGGAGGCCCGGCGAGTGTCTTCTCAGACGGAATGCAGCTCAAGAAGTACCAGGGCATGGGCTC CTGGATGCCATGGAGAAGAGCAGCAGCAGCCAGAAACAATACTTCAACGACGGGGATAAGGCGAAGATC ACGCAGGATGTCTGGGCTCCATCCAGGACAAAGGGTCCATTGAGAAGTTCGTGCCTACCTCATAGTGG GCATCCAGCATGGCTGCCAGGATATCGGGGCCACAGCCTGTCTGTCTTCGGTCCATGATGTACTCAGG GGAGTCAAGTTTGAGAAGCAGACCATGTGAGCCAGATCGACGGTGGCATCCATGGCCTGCACTCTTAC GAGAAGTGGCTGTACTGAGGACAGCGGTGCAGGGCGAGATG

A NOV93 polypeptide (SEQ ID NO:296) encoded by SEQ ID NO:295 is 518 amino acids in length and is presented using the one-letter amino acid code in Table 93B. The Psort profile for NOV93 predicts that this sequence is likely to be localized to the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV93 polypeptide is located to lysosomes with a certainty of 0.1921, or, to microbodies with a certainty of 0.3346.

Table 93B. NOV93 Polypeptide Sequence (SEQ ID NO:296)

```
MRGLPISSSTGYVVEDGFTATVQQLFASDQGLTYNDFLILPGFIDFIADVDLTSLTQTP
VTLKTPLISSPMDTVTEADLAIVMALMGGTGFIHHNCTPEFQASEVQKVKKFEPGFITHP
VVLSPHPTVGDVLEAKMRHGFSGIPITETGTMGSKLVGIVTSRDIDFLAEKDHTLLSEV
MMPRIKLVVAPASSVRLKEANEILQLSKKGKLPVNDRELVAIITCTALKNRDYPVASK
DSHEQLLGGAAGVTHEDDKYHLDLLTQVGVNVIGLDSSQGNVYQIAMVHYIKQKYPHLQ
VIGGNVVTAAQANNLIDAGVDGLGRGMDCAAGSIYINQEVIACSQPQGTAVYKVAKHTQN
FGVPIIADGGIQTMGHVVKALALGASTVMMGSLAATMEAPGECFFSDGMQLKKYQGMGS
LDAMEKSSSSQKQYFNDGDKAKITQDVLGSIQDKGSIQKFVYPYLIVGIQHGCDIGAHSL
SVLRSMYSGELKFEKQTMSAQIDGGIHLHSYEKWLY
```

A BLAST analysis of NOV93 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV93 had high homology to other proteins as shown in Table 93C.

Table 93C. BLASTX results from PatP database for NOV93

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAR05431 Chinese hamster IMPDH - <i>Cricetulus</i> sp, 514 aa.	1874	3.2e-193
patp:AAR05432 Human IMPDH - <i>Homo sapiens</i> , 514 aa.	1872	5.3e-193
patp:AAY08965 <i>A. gossypii</i> inosine-monophosphate dehydrogenase	980	1.8e-98
patp:AAG30888 <i>Arabidopsis thaliana</i> protein fragment	973	9.7e-98
patp:AAG43108 <i>Arabidopsis thaliana</i> protein fragment	949	3.4e-95

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1402 of 1567 bases (89%) identical to a gb:GENBANK-ID:HUMIMPH|acc:J05272.1 mRNA from *Homo sapiens* (Human IMP dehydrogenase type 1 mRNA). The full amino acid sequence of the protein of the invention was found to have 438 of 513 amino acid residues (85%) identical to, and 462 of 513 amino acid residues (90%) similar to, the 514 amino acid residue ptnr:SWISSNEW-ACC:P20839 protein from *Homo sapiens* (Human) (INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE 1 (EC 1.1.1.205) (IMP DEHYDROGENASE 1) (IMPDH-I) (IMPD 1)). NOV93 also has homology to the other proteins shown in the BLASTP data in Table 93D.

Table 93D. NOV93 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17453301 ref XP_004627.5 (XM_004627)	IMP (inosine monophosphate) dehydrogenase 1 [Homo sapiens]	514	440/514 (85)	464/514 (89)	0.0
gi 124417 sp P20839 IMD1_HUMAN	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE 1 (IMP DEHYDROGENASE 1) (IMPDH-I) (IMPD 1)	514	438/514 (85)	462/514 (89)	0.0
gi 4504687 ref NP_000874.1 (NM_000883)	IMP (inosine monophosphate) dehydrogenase 1; SWSS2608 [Homo sapiens]	514	434/514 (84)	458/514 (88)	0.0
gi 6754344 ref NP_035959.1 (NM_011829)	inosine 5'-phosphate dehydrogenase 1 [Mus musculus]	514	430/514 (83)	462/514 (89)	0.0
gi 16549223 dbj BAB70780.1 (AK054640)	unnamed protein product [Homo sapiens]	489	418/514 (81)	440/514 (85)	0.0

5 This BLASTP data is displayed graphically in the ClustalW in Table 93E. A multiple sequence alignment is given, with the NOV93 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 93D.

Table 93E. ClustalW Alignment of NOV93	
NOV93 (SEQ ID NO:296)	
gi 17453301 (SEQ ID NO:764)	
gi 124417 (SEQ ID NO:765)	
gi 4504687 (SEQ ID NO:766)	
gi 6754344 (SEQ ID NO:767)	
gi 16549223 (SEQ ID NO:768)	
	10 20 30 40 50
NOV93	MRGLPISSTGYVVEDGEFTATVQQLFASDQGLTYNDFLILPGFIDFIADE
gi 17453301	MADYLISGGTGYVPEDGLTA--QQLFASADGLTYNDFLILPGFIDFIADE
gi 124417	MADYLISGGTGYVPEDGLTA--QQLFASADGLTYNDFLILPGFIDFIADE
gi 4504687	MADYLISGGTGYVPEDGLTA--QQLFASADGLTYNDFLILPGFIDFIADE
gi 6754344	MADYLISGGTGYVPEDGLTA--HQLFANADGLTYNDFLILPGFIDFIADE
gi 16549223	MADYLISGGTGYVPEDGLTA--QQLFASADGLTYNDFLILPGFIDFIADE
	60 70 80 90 100
NOV93	VDLTSALTQPTLTKTPLISSPMDTVTEADLAI VMALMGCTGFIHNCNCTPE
gi 17453301	VDLTSALTRKITLTKTPLISSPMDTVTEADMAIAMALMGGIGFIHNCNCTPE
gi 124417	VDLTSALTRKITLTKTPLISSPMDTVTEADMAIAMALMGGIGFIHNCNCTPE
gi 4504687	VDLTSALTRKITLTKTPLISSPMDTVTEADMAIAMALMGGIGFIHNCNCTPE
gi 6754344	VDLTSALTRKITLTKTPLISSPMDTVTEADMAIAMALMGGIGFIHNCNCTPE
gi 16549223	VDLTSALTRKITLTKTPLISSPMDTVTEADMAIAMA-----
	110 120 130 140 150
NOV93	FOASEVQKVKKEFGGFITHPVVLSPLHTVGDVLEAKMRHGFSGIPITETG

gi 17453301	FOANEVRKVKKFEQGFI	TDPVVLSPSHTVGDVLEAKMRHGFSGIPITETG	
gi 124417	FOANEVRKVKNF	EQGFI	TDPVVLSPSHTVGDVLEAKMRHGFSGIPITETG
gi 4504687	FOANEVRKVKNF	EQGFI	TDPVVLSPSHTVGDVLEAKMRHGFSGIPITETG
gi 6754344	FOANEVRKVKKFEQGFI	TDPVVLSPSHTVGDVLEAKMRHGFSGIPITATG	
gi 16549223	-----KFEQGFI	TDPVVLSPSHTVGDVLEAKMRHGFSGIPITETG	

	160	170	180	190	200
NOV93	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	MPRIKLVVAPASSVRLKEA		
gi 17453301	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	TPRIELVVAPAG-VTLKEA		
gi 124417	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	TPRIELVVAPAG-VTLKEA		
gi 4504687	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	TPRIELVVAPAG-VTLKEA		
gi 6754344	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	TPRIELVVAPAG-VTLKEA		
gi 16549223	TMGSKLVGI	VTSRDIDFLAEKDHTTLLSEVM	TPRIELVVAPAG-VTLKEA		

	210	220	230	240	250
NOV93	NEILQSKKGLPIVND	DELVAII	CTALK-NRDYPVASKDSHEQLLGG		
gi 17453301	NEILQSKKGLPIVND	CDELVAII	ARTDLKKNRDYPLASKDSQKQLLGG		
gi 124417	NEILQSKKGLPIVND	CDELVAII	ARTDLKKNRDYPLASKDSQKQLLGG		
gi 4504687	NEILQSKKGLPIVND	CDELVAII	ARTDLKKNRDYPLASKDSQKQLLGG		
gi 6754344	NEILQSKKGLPIVND	CDELVAII	ARTDLKKNRDYPLASKDSHQLLGG		
gi 16549223	NEILQSKKGLPIVND	CDELVAII	ARTDLKKNRDYPLASKDSQKQLLGG		

	260	270	280	290	300
NOV93	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				
gi 17453301	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				
gi 124417	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				
gi 4504687	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				
gi 6754344	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				
gi 16549223	AAVGTREDDKYRLDLLTQAGVDVIVLDSSQGNSVYQIAMVHYIKQKYPHL				

	310	320	330	340	350
NOV93	QVIGGNVVTAQAANLIDAGVDGLGRGMDCAAGSTYINQEVHACSPQGT				
gi 17453301	QVIGGNVVTAQAANLIDAGVDGLRVGMGC--GSICITQEVMACGRPQGT				
gi 124417	QVIGGNVVTAQAANLIDAGVDGLRVGMGC--GSICITQEVMACGRPQGT				
gi 4504687	QVIGGNVVTAQAANLIDAGVDGLRVGMGC--GSICITQEVMACGRPQGT				
gi 6754344	QVIGGNVVTAQAANLIDAGVDGLRVGMGC--GSICITQEVMACGRPQGT				
gi 16549223	QVIGGNVVTAQAANLIDAGVDGLRVGMGC--GSICITQEVMACGRPQGT				

	360	370	380	390	400
NOV93	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			
gi 17453301	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			
gi 124417	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			
gi 4504687	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			
gi 6754344	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			
gi 16549223	AVYKVAEYARRFGVPPIADGGIQT	VGHVVKALALGASTVMMGSLLAATTE			

	410	420	430	440	450
NOV93	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				
gi 17453301	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				
gi 124417	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				
gi 4504687	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				
gi 6754344	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				
gi 16549223	APGEYFFSDGVLKLYRGMSLDAMEKSSSSQKRYFSEGDVKVIAQGVSG				

	460	470	480	490	500
NOV93	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			
gi 17453301	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			
gi 124417	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			
gi 4504687	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			
gi 6754344	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			
gi 16549223	SIQDKGSIQKFVPLYIAGIQHCQDIGA	SLSVLRSMMSYSGELKFEKRTM			

		510
NOV93	SAQIDGGTHGLHSYEKWL	
gi 17453301	SAQIEGGVHGLHSYEKRL	
gi 124417	SAQIEGGVHGLHSYEKRL	
gi 4504687	SAQIEGGVHGLHSYEKRL	
gi 6754344	SAQIEGGVHGLHSYEKRL	
gi 16549223	SAQIEGGVHGLHSYEKRL	

Table 93F lists the domain description from DOMAIN analysis results against NOV93. This indicates that the NOV93 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 93F. Domain Analysis of NOV93				
gnl Pfam pfam00478, IMPDH_C, IMP dehydrogenase / GMP reductase C terminus. This family is involved in biosynthesis of guanosine nucleotide biosynthesis. Members of this family contain a TIM barrel structure. The alignment does not contain the whole TIM barrel domain. The alignment is truncated after the insert domain (2 CBS domains pfam00571) found in the inosine-5'-monophosphate dehydrogenase structure. This family should always be associated with pfam01574. This family is a member of the common phosphate binding site TIM barrel family. SEQ ID NO:884				
CD-Length = 222 residues, 100.0% aligned Score = 261 bits (666), Expect = 9e-71				
NOV93:	264	LLTQVG VNVIGLDSSQ GNSVYQIAMVHYIKQKYPHLQVIGNVVTAAQANNLIDAGVDGL	323	
		L + GV+VI LDSS G S QI + +I++KYP +QVI GNVVT A LIDAG D +		
Sbjct:	1	ALVEAGVDVICLDSSNGYSEVQIDFIRWIREKYPTVQVIAGNVVTGEMAEELIDAGADAV	60	
NOV93:	324	GRGMDCAAGSIYINQEVIA CSQPQGTAVYKVAKHTQNFQVPIIADGGIQTMGHVVKALAL	383	
		G+ GSI I +EV +PQ TAV +VA + +P+I+DGGI GH+ KALA		
Sbjct:	61	KVGI--GPGSICITREVAGIGRPQATAVLEVADASHGLNIPVISDGGITNPGHMAKALAG	118	
NOV93:	384	GASTVMMGSLLAATMEAPGECFFSDGMQLKKYQGMGSLDAMEKSSSSQKQYFNDGDKAKI	443	
		GA VM+GSLLA T EAPGE F DG + K Y+GMGSL AM+K S +YF K +		
Sbjct:	119	GADFVMIGSLLAGTTEEAPGEVVFQDKGKKYKLYRGMGSLTAMKKYQGSVARYFASKQKLSV	178	
NOV93:	444	TQDVLGSIQDKGSIQKFVPLYLIVGIQHGCQDIGAHSLSVLRSM	487	
		+ V G + KG + + V L+ G++ C IGA L LR		
Sbjct:	179	EEGVTGYVPYKGDVSRVTVDLLGGLRSSCTYIGATKLKQLRKRA	222	

Inosine-5-prime-monophosphate dehydrogenase (EC 1.1.1.205) catalyzes the formation of xanthine monophosphate (XMP) from IMP. In the purine de novo synthetic pathway, IMP dehydrogenase is positioned at the branch point in the synthesis of adenine and guanine nucleotides and is thus the rate-limiting enzyme in the de novo synthesis of guanine nucleotides. Inhibition of cellular IMP dehydrogenase activity results in an abrupt cessation of DNA synthesis and a cell-cycle block at the G1-S interface. Collart and Huberman (1988) used a polyclonal antibody directed against the purified protein to isolate human and Chinese hamster IMP dehydrogenase cDNA clones. The sequence of these clones demonstrated an open reading frame for a protein containing 514 amino acids. The molecular mass of the

15

produced protein was 56 kD, which is the observed molecular mass of the purified protein and of the immunoprecipitated in vitro translation product.

A high order of conservation of the IMP dehydrogenase protein was indicated by the finding that human and Chinese hamster cDNA clones differed by only 8 amino acids.

5 Natsumeda et al. (1990) isolated two distinct cDNAs (types I and II) encoding IMP dehydrogenase from a human spleen cDNA library. Both clones encode proteins of 514 residues showing 84% sequence identity. Type I mRNA was found to be the main species in normal leukocytes, and type II (146691) predominated in human ovarian tumors. Using PCR primers specific for type II IMPDH, Glesne et al. (1993) screened a panel of human/Chinese
10 hamster cell somatic hybrids and a separate deletion panel of chromosome 3 hybrids and localized the gene to 3p24.2-p21.2.

The gene was also localized on a map of two overlapping YACs and found to span no more than 12.5 kb of genomic DNA. From cloning and sequencing IMPD2, Glesne and Huberman (1994) determined that the gene spans approximately 5 kb and is interrupted by 12
15 introns. The transcriptional start sites were determined by S1 nuclease mapping to be somewhat heterogeneous but the predominant mRNA species showed a 5-prime end at 102 and 85 nucleotides from the translational initiation codon. Zimmermann et al. (1995) also cloned the human gene and noted that it has 14 exons spanning approximately 5.8 kb. They also characterized regulatory elements in the 5-prime flanking region of the gene.

20 NOV93 is predicted to be expressed in at least the following tissues: brain, prosencephalon/forebrain, diencephalon, pituitary gland. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV93 is provided in Example 2.

25 The NOV93 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral
30 disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV93 nucleic acids encoding the inosine-5-prime-monophosphate dehydrogenase-like protein of the invention, or fragments thereof, may further be useful in

diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a inosine-5'-monophosphate dehydrogenase-like protein includes the nucleic acid whose sequence is provided in Table 93A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 93A while still encoding a protein that maintains its inosine-5'-monophosphate dehydrogenase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 93A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 11% of the bases may be so changed.

The novel protein of the invention includes the inosine-5'-monophosphate dehydrogenase-like protein whose sequence is provided in Table 93B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 93B while still encoding a protein that maintains its inosine-5'-monophosphate dehydrogenase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 15% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV94

The disclosed NOV94 (alternatively referred to herein as CG56281-01) includes the 1573 nucleotide sequence (SEQ ID NO:297) shown in Table 94A. A NOV94 ORF begins

with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 1564-1566. The disclosed NOV94 maps to human chromosome 3.

Table 94A. NOV94 Nucleotide Sequence (SEQ ID NO:297)

```
ATGGCGCGGAAGCAGGACCCGAAGCCTAAATTCAGGAGGGTGAGCGAGTGCTGTGCTTTCATGGGCCTC
TTCTTTATGAAGCAAAGTGTGTAAAGTTGCCATAAAGGACAAACAAGTGAAATACCTTCATACATTACAG
TGGTTTGGAAATAAAATTTGGGATGAGTGGGTTCCGGAGAGCAGAGTACTCAAATACGTGGACACCAATGAA
AATCGTAGATTGGCCAGGGAAATTCGTAGATTACAGCATAAATTGGCAAGAAATGCTGTAGCTCACCTGA
GGAGCAAGAGAGAAAGAAGCAGCCGCTCCAGGTTGCTTGGTGCTGACTCTGTCTTAAAGGCCTCTCCAT
CGAAGAAAAAATGAAAATGATGAAAACCTCATTAAAGCAGTTCTCTGACAGTAGTGAAGACAAAGGATGAA
AAAATAAGTGAAGAATGTGATATTGAAGAAAAGACTGAAGTGAAAGAAGAACCGGAGCTTCAAACAAAAA
GGGAAATGGAAGAAAGAACAGTAACCTTAGAAATCCCTGAAGTTCTGAAGAGGCAGCTGGAGGATGATTG
TTACTACATTAATCGGAGGAAACGGTTAGTGCAACTTCCATGCCACACCAACATCATAACGATTTTGGAA
TCCTATGTGAAGCATTTTGTCTATCAGTGCAGCCTTTTCAGCCAATGAGAGGCCCTCGTCACCATCACGCTA
TGCCACATGCCAGCATGAACGTGCCTTATATCCCAGCAGAAAAGAATATTGACCTTTGTAAGGAGATGGT
GGATGGATTAAAGAATAACCTTTGATTACACTCTCCGTTGGTTTACTCTATCCCTATGAACAAGCTCAG
TATAAAAAGGTGACTGCATCTAAGGTTTCTTGCAATTAAGGAAAGTGCCACAAATACTAATAGGAGCC
AGGAGAAGCTCTCTCCAGCTTACGTTTGTGTAATCCATCCAGGCCGAGTCTACAGAGAGTCAGTCGAC
CAGCGGTGAACCAAGCCACCCCTAAAAGGCGCAAAGCCGAGCCGCAAGCAGTGCAGTCTCTGAGGCGGTCC
TCGCCCCACACCGCAACTGTGACAGGCTTTCTAAGAGCAGCACCTCACCTCAGCCCAAGCGCTGGCAGC
AGGACATGTCCACCAGTGTGCCAAGCTTCTCTGCACCTGGAAAAGAAGACACCTGTGCATAGCAGATC
ATCTTCACCTACTCTGACTCCTAGCCAGGAAGGGAGTCTGTGTTTGTCTGGCTTTGAAGGGAGAAGAACT
AATGAAATAAATGAGGTCCTCTCTGGAAGCTCGTACCTGACAATTACCCACCAGGTGACCAGCCACCTC
CACCTCTTACATTTACGGGGCGCAACATTTGCTGCGATTGTTGTCAAACCTCCAGAAATCTTGGAAA
AATGTCTTTACTGAGAAGAATCTGAAGGCTTTATTGAAGCACTTTGATCTCTTTGTGAGGTTTGTAGCA
GAATACCACGATGACTTCTTCCCAGAGTCAGCTTACGTGCTGCCTCTGAGGTGCATTACAGCACCAGGA
ACCCCGAGGCAGTCAATAAGTGTGATGGTTCT
```

A NOV94 polypeptide (SEQ ID NO:298) encoded by SEQ ID NO:297 is 521 amino acids in length and is presented using the one-letter amino acid code in Table 94B. The Psort profile for NOV94 predicts that this sequence is likely to be localized to the nucleus with a certainty of 0.9700. In alternative embodiments, a NOV94 polypeptide is localized to microbodies with a certainty of 0.3000.

Table 94B. NOV94 Polypeptide Sequence (SEQ ID NO:298)

```
MARKQDPKPKFQEGERVLCFHGPLYEAKCVKVAIKDKQVKYFIHYSWGNKNWDEWPES
RVLKYVDTNENRRLAREIRRLQHKLARNVAHLRSKRERSRRSLRGADSVLKGLSIEEK
NENDENSLSSSSDSEDKDEKISEECDIEEKTEVKEEPELQTKREMEERTVTLEIPEVLK
RQLEDDCYIINRRKRLVQLPCHTNIITILESIVKHFAISAAFSANERPRHHAMPHASMN
VPYI PAEKNIDLCKEMVDGLRITFDYTLPLVLLYPYEQAYKKVTSKVF LAIKESATNT
NRSQEKLSPSLRLLNPSRPQSTESQSTSGEPATPKRRKAEPQAVQSLRRSSPHANTCDRL
SKSSTSPQPKRWQDMSTSVPKLFLHLEKTPVHSRSSPTLTPSQEGSPVFAGFEGRRRT
NEINEVLSWKLVPDNYPPGDQPPPSYIYGAQHLLRLFVKLPEILGKMSFTEKNLKALLK
HFDLFVRFLAEYHDDFFPESAYVAASEVHYSTRNPQAVNKC
```

A BLAST analysis of NOV94 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV94 had high homology to other proteins as shown in Table 94C.

Table 94C. BLASTX results from PatP database for NOV94

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)
patp:AAW96184 Senescence protein derived from human	465	2.9e-51
patp:AAB60085 Human transport protein TPPT-5	465	2.9e-51
patp:AAU32295 Novel human secreted protein #2786	430	3.4e-40
patp:AAM64801 Human brain expressed single exon probe	239	3.5e-19
patp:AAM77558 Human bone marrow expressed probe	239	3.5e-19

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1381 of 1581 bases (87%) identical to a gb:GENBANK-ID:AF117065|acc:AF117065.1 mRNA from *Homo sapiens* (male-specific lethal-3 homolog 1 (MSL3L1) mRNA). The full amino acid sequence of the protein of the invention was found to have 414 of 520 amino acid residues (79%) identical to, and 457 of 520 amino acid residues (87%) similar to, the 521 amino acid residue ptnr:SPTREMBL-ACC:Q9Y5Z8 protein from *Homo sapiens* (Human) (MALE-SPECIFIC LETHAL-3 HOMOLOG 1). NOV94 also has homology to the other proteins shown in the BLASTP data in Table 94D.

Table 94D. NOV94 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17975757 ref NP_523353.1 (NM_078629)	male-specific lethal 3-like 1 isoform a; drosophila MSL3-like 1 [<i>Homo sapiens</i>]	521	413/529 (78)	456/529 (86)	0.0
gi 17975759 ref NP_523354.1 (NM_078630)	male-specific lethal 3-like 1 isoform b; drosophila MSL3-like 1 [<i>Homo sapiens</i>]	462	389/468 (83)	420/468 (89)	0.0
gi 14764458 ref XP_045715.1 (XM_045715)	male-specific lethal-3 (Drosophila)-like 1 [<i>Homo sapiens</i>]	496	387/498 (77)	427/498 (85)	0.0
gi 11545735 ref NP_034962.2 (NM_010832)	male-specific lethal-3 homolog 1 (Drosophila) [<i>Mus musculus</i>]	466	346/472 (73)	396/472 (83)	e-168
gi 17975761 ref NP_006791.2 (NM_006800)	male-specific lethal 3-like 1 isoform c; drosophila MSL3-like 1 [<i>Homo sapiens</i>]	355	312/354 (88)	335/354 (94)	e-162

This BLASTP data is displayed graphically in the ClustalW in Table 94E. A multiple sequence alignment is given, with the NOV94 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 94D.

Table 94E. ClustalW Alignment of NOV94

NOV94	(SEQ ID NO:298)
gi 17975757	(SEQ ID NO:769)
gi 17975759	(SEQ ID NO:770)
gi 14764458	(SEQ ID NO:771)
gi 11545735	(SEQ ID NO:772)
gi 17975761	(SEQ ID NO:773)

	10	20	30	40	50
NOV94	MARKQDPKPKFQEGERVLCFHG	-----	PLLYEAKCVKVAIK-D	-----	KQV
gi 17975757	MSASEGMKFKFHSGEKVLCFEPDPTKARVLYDAKIVVVIVGKDEKGRKIP				
gi 17975759	-----				
gi 14764458	MSASEGMKFKFHSGEKVLCFEPDPTKARVLYDAKIVDVIVGKDEKGRKIP				
gi 11545735	-----				
gi 17975761	-----				

	60	70	80	90	100
NOV94	KYFIHYSGWKNWDEWVPE	SRVLKYVD	TNENRRLAREIR	RLQHKLARN	NAV
gi 17975757	EYLIHFNGWNR	SWDRWAAEDHVL	R-----	DTDENR	-----
gi 17975759	-----	MPSWDRWAAEDHVL	R-----	DTDENR	-----
gi 14764458	EYLIHFNGWNR	SWDRWAAEDHVL	R-----	DTDENR	-----
gi 11545735	-----	MPSWDRWAAEDHVL	R-----	DTDENR	-----
gi 17975761	-----	MPSWDRWAAEDHVL	R-----	DTDENR	-----

	110	120	130	140	150
NOV94	AHLRSKRERSS	SRLLGADSVLKGL	STEEKNNENDENSLSSSS	DSSEDK	
gi 17975757	ARLRSTGRKKKRCRLPGVDSVLKGLPTEEKDENDENSLSSSS	DCSENKD			
gi 17975759	ARLRSTGRKKKRCRLPGVDSVLKGLPTEEKDENDENSLSSSS	DCSENKD			
gi 14764458	ARLRSTGRKKKRCRLPGVDSVLKGLPTEEKDENDENSLSSSS	DCSENKD			
gi 11545735	ARLRSTGRKKKRCRLPGVDSVLKGLPTEEKDENDENSLSSSS	DCSENKD			
gi 17975761	ARLRSTGRKKKRCRLPGVDSVLKGLPTEEKDENDENSLSSSS	DCSENKD			

	160	170	180	190	200
NOV94	EKISE	ESDIEEKTEVKEEP	ELQTRREMEERTITIEIPEVLKKQLEDD		
gi 17975757	EEISE	ESDIEEKTEVKEEP	ELQTRREMEERTITIEIPEVLKKQLEDD		
gi 17975759	EEISE	ESDIEEKTEVKEEP	ELQTRREMEERTITIEIPEVLKKQLEDD		
gi 14764458	EEISE	ESDIEEKTEVKEEP	ELQTRREMEERTITIEIPEVLKKQLEDD		
gi 11545735	GGIKRHRORRIKVKAKAKKVL	SLRSRKEMDERTITIDIP	VLKKQLEDD		
gi 17975761	-----	MEERTITIEIPEVLKKQLEDD			

	210	220	230	240	250
NOV94	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH
gi 17975757	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH
gi 17975759	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH
gi 14764458	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH
gi 11545735	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH
gi 17975761	CYYINRRKRLVQLPCH	TNIITILES	YVKHF	FAISAAFSANERPRHHH	AMPH

	260	270	280	290	300
NOV94	ASMNVPYI	PAEKNIDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTA	
gi 17975757	ANMNVHYI	PAEKNVDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTS	
gi 17975759	ANMNVHYI	PAEKNVDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTS	
gi 14764458	ANMNVHYI	PAEKNVDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTS	
gi 11545735	ANMNVHYI	PAEKNVDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTS	
gi 17975761	ANMNVHYI	PAEKNVDLC	KEMVDGLRITFDYTLPLVLLYPYEQ	AQYKKVTS	

	310	320	330	340	350
NOV94	SKVFLAIKESATNTNRSQ	EKLSPSLRL	LLNPSRPOSTESQ	STSGEPATPKR	
gi 17975757	SKFFLP	IKESATNTNRSQ	EELSPSP	LLNPSTPOSTESQ	PTTGEPATPKR
gi 17975759	SKFFLP	IKESATNTNRSQ	EELSPSP	LLNPSTPOSTESQ	PTTGEPATPKR

gi	14764458	SKFFFLPIKESATSTNRSQEELSPSPPLLNPSTPQSTESQPTTGEPATPKR
gi	11545735	SKFFFLPIKESATSTNRSQEELSPSPPLLNPSTPQSTESQPTTGEPATPKR
gi	17975761	SKFFFLPIKESATSTNRSQEELSPSPPLLNPSTPQSTESQPTTGEPATPKR
<div> <div>360370380390400</div> <div> <div>NOV94</div> <div>gi 17975757</div> <div>gi 17975759</div> <div>gi 14764458</div> <div>gi 11545735</div> <div>gi 17975761</div> </div> </div>		
<div> <div>410420430440450</div> <div> <div>NOV94</div> <div>gi 17975757</div> <div>gi 17975759</div> <div>gi 14764458</div> <div>gi 11545735</div> <div>gi 17975761</div> </div> </div>		
<div> <div>460470480490500</div> <div> <div>NOV94</div> <div>gi 17975757</div> <div>gi 17975759</div> <div>gi 14764458</div> <div>gi 11545735</div> <div>gi 17975761</div> </div> </div>		
<div> <div>510520530</div> <div> <div>NOV94</div> <div>gi 17975757</div> <div>gi 17975759</div> <div>gi 14764458</div> <div>gi 11545735</div> <div>gi 17975761</div> </div> </div>		

Table 94F lists the domain description from DOMAIN analysis results against NOV94. This indicates that the NOV94 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 94F. Domain Analysis of NOV94				
gn Smart smart00298, CHROMO, Chromatin organization modifier domain SEQ ID NO: 885				
CD-Length = 55 residues, Score = 42.0 bits (97), Expect = 9e-05				
NOV94:	36	KDKQVKYFIHYSGWKNWDEWVPESRVLK	64	
		K +++Y + + G++ D W PE +L		
Sbjct:	14	KKGELEYLVKWKGYSYREDTWEPEENLLN	42	

The Drosophila male-specific lethal (msl) genes regulate transcription from the male X chromosome in a dosage compensation pathway that equalizes X-linked gene expression in males and females. The members of this gene family, including msl1, msl2, msl3, mle, and mof, encode proteins with no sequence similarity to known proteins. However, mutations in

10

each of these genes produce a similar phenotype: sex-specific lethality of male embryos caused by the failure of mutants to increase transcription from the single male X chromosome.

The MSL gene products assemble into a multiprotein transcriptional activation complex at hundreds of sites along the chromatin of the X chromosome. By searching
5 sequence databases with the sequence of a BAC clone that maps to Xp22.3, Prakash et al. (1999) identified a human homolog of *Drosophila* msl3, MSL3-like-1 (MSL3L1). They isolated a cDNA containing a complete MSL3L1 coding sequence. The deduced 521-amino acid MSL3L1 protein shares 30% overall sequence identity with *Drosophila* MSL3 and 86% identity with mouse Msl3l1. Three segments of the *Drosophila* MSL3 protein are highly
10 conserved in MSL3L1, including two putative chromodomains, one at the N terminus and the other at the C terminus. Chromodomains, which form a characteristic tertiary structure and can interact with components of chromatin, have been implicated to play roles in chromatin organization and transcriptional regulation. MSL3L1 also contains a putative nuclear localization signal, a putative leucine zipper motif within the second chromodomain, and two
15 potential tyrosine kinase phosphorylation sites.

Prakash et al. (1999) identified human fetal kidney cDNAs representing an alternatively spliced MSL3L1 transcript that lacks exon 2. The predicted protein, which is referred to as isoform 2, is identical to the first isoform from amino acid 62 to the C terminus but does not contain the first 26 amino acids of the N-terminal chromodomain. Northern blot
20 analysis detected a major 2.4-kb MSL3L1 transcript in all tissues examined, namely liver, pancreas, heart, lung, kidney, skeletal muscle, brain, and placenta, with highest expression in skeletal muscle and heart. A 2.6-kb transcript unique to skeletal muscle was also found. Northern blot analysis of E7, E11, E15, and E17 mouse embryos detected approximately equal levels of Msl3l1 expression in all embryos. The MSL3L1 gene spans 17 kb and contains 13
25 exons. It is transcribed from telomere to centromere. Prakash et al. (1999) showed that the MSL3L1 gene undergoes X inactivation.

NOV94 is predicted to be expressed in at least the following tissues: lung, testis, B-cell. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST
30 sources, literature sources, and/or RACE sources. Further expression data for NOV94 is provided in Example 2.

The NOV94 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis,

hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation,
5 immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV94 nucleic acids encoding the MSL3L1-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

10 The novel nucleic acid of the invention encoding a male-specific lethal 3-like 1-like protein includes the nucleic acid whose sequence is provided in Table 94A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 94A while still encoding a protein that maintains its male-specific lethal 3-like 1-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences
15 are complementary to the sequence of Table 94A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar
20 phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 13% of the bases may be so changed.

25 The novel protein of the invention includes the male-specific lethal 3-like 1-like protein whose sequence is provided in Table 94B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 94B while still encoding a protein that maintains its male-specific lethal 3-like 1-like activities and physiological functions, or a functional fragment thereof. In the mutant or
30 variant protein, up to about 21% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using

prediction from hydrophobicity charts, as described in the “Anti-NOVX Antibodies” section below.

NOV95

- 5 The disclosed NOV95 (alternatively referred to herein as CG56975-01) includes the 1323 nucleotide sequence (SEQ ID NO:299) shown in Table 95A. A NOV95 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 35-37 and ends with a stop codon at nucleotides 1301-1303. The disclosed NOV95 maps to human chromosome 1.

Table 95A. NOV95 Nucleotide Sequence (SEQ ID NO:299)

GCAATTCCTTTTCAATATTTATATATTTTCAGAAAATGTCAGTAAATTCACAAATGCAAAACGGATTGAA
GGACTTGATAGTAATGTGTGGATTGAATTTACCAAATTTGGCTGCAGACCCTTCTGTTGTGAATCTTGGCC
AAGGCTTTCCAGATATATCCCTCCTACATATGTAAGAAGAATTATCAAAGATTGCAGCAATCGATAG
CCTGAATCAGTATACACGAGGCTTTGGCCATCCATCACTTGTGAAAGCTCTGTCTATCTGTATGAAAAG
CTTTATCAAAGCAAATTGATTCAAATAAAGAATCCTTGTGACAGTAGGAGCATATGGATCTCTTTTAA
ACACCATTCAGCATTAATTGATGAGGGACAGGTCACTAATAGTGCCTTTCTATGACTGCTATGAGCC
CATGGTGAGAATGGCTGGAGCAACACCTGTTTTTATCCCTGAGATCTGTAAGTTTGGGAAAAGATGG
TCTAGTTCTGACTGGACATTAGATCCTCAAGAACTGGAAAGTAAATTTAATTCAAAACCAAAGCTATTA
TACTAAATACTCCACATAACCCACTTGGCAAGGTATATAACAGAGAGGAAGTCAAGTAATTGCTGACCT
TTGCATCAAATATGACACACTCTGCATCAGCGATGAGGTTTATGAATGGCTTGTATATTCTGGAATAAG
CACTTAAAAATAGCTACTTTTCCAGGTATGTGGGAGAGACAATAACAATAGGAAGTCTGGAAAGACTT
TCAGTGTAAGTGGCTGGAAGGTAGGCTGGTCCATTTGGTCCAAATCATTGATAAAACATTTACAGACAGT
TCAACAAAACACGATTTATACTTGTGCACTCCTTTACAGGAAGCCTTGGCTCAAGCTTTCTGGATTGAC
ATCAAGCGCATGGATGACCCAGAATGTTACTTTAATCTTTGCCAAAAGAGTTAGAAGTAAAAGAGATC
GGATGGTACGTTTACTTGAAAGTGTGGCCTAAAACCCATAGTTCTGATGGAGGATACTTCATCATCGC
TGATGTGTCTATTTTCAATTGTGGTTTTAGATCCAGACCTCTCTGATATGAAGAATAATGAGCCTTATGAC
TATAAGTTTGTGAAATGGATGACTAAACATCAGAACTATCAGCCATCCCCGTTTTCAGCATTCTGTAAT
CAGAGACTAAATCACAGTTTGAAGGTTTGTGCGTTTGTGCTTCATTAAAGTAAGTTCCCTGCTCGATGC
TGCTGAAGAAATCATCAAGGCATGGAGTGTACAGAAGTCTTGATTGTGACAGATGGATTAAT

- 10 A NOV95 polypeptide (SEQ ID NO:300) encoded by SEQ ID NO:299 is 422 amino acids in length and is presented using the one-letter amino acid code in Table 95B. The Psort profile for NOV95 predicts that this sequence is a Type II membrane protein, and is likely to be localized at the plasma membrane with a certainty of 0.4400. In alternative embodiments, a
15 NOV95 polypeptide is located to microbodies with a certainty of 0.3691.

Table 95B. NOV95 Polypeptide Sequence (SEQ ID NO:300)

MSLKFTNAKRIEGLDSNVWIEFTKLAADPSVVNLGQGFDPISPTTYVKEELSKIAAIDSL
NQYTRGFHPSLVKALSYLEKLYQKQIDSNKEILVTVGAYGSLFNTIQALIDEGQVILI
VPFYDCYEPVMRMAGATPVFIPLRSVSLGKRWSDDWTLDPQELESKFNSKTKAILNTP
HNPLGKVYNREELQVIADLCIKYDTLCISDEVYEWLVYSGNKHKLKIATFPGMWERTITIG
SAGKTFSVTGKVGWSIGPNHLIKHLQTVQONTIYTCATPLQEAQAQAFWIDIKRMDDE
CYFNSLPKELEVKRDRMVRLLESVGLKPIVPDGGYFIIADVSIFIVLDPDLSDMKNNP
YDYKFVKWMTKHQKLSAIPVSAFCNSETKSQFEKFVRFCEIKVSSLLDAAEEIIKAWSVQ
KS

A BLAST analysis of NOV95 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV95 had high homology to other proteins as shown in Table 95C.

Table 95C. BLASTX results from PatP database for NOV95

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)
patp:AAV54591 Amino acid sequence of a human transferase	2121	2.2e-219
patp:AAR89906 Human kynurenine aminotransferase (KAT)	1171	1.0e-118
patp:AAR89896 Rat kynurenine aminotransferase (KAT)	1130	2.2e-114
patp:AAR89897 Rat kynurenine aminotransferase (KAT)	1130	2.2e-114
patp:AAR89898 Rat kynurenine aminotransferase (KAT)	1130	2.2e-114

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 806 of 915 bases (88%) identical to a gb:GENBANK-ID:AF091090|acc:AF091090.1 mRNA from *Homo sapiens* (clone 669 unknown mRNA, complete sequence). The full amino acid sequence of the protein of the invention was found to have 229 of 415 amino acid residues (55%) identical to, and 300 of 415 amino acid residues (72%) similar to, the 419 amino acid residue ptrn:SPTREMBL-ACC:Q9W6U2 protein from *Fugu rubripes* (Japanese pufferfish) (*Takifugu rubripes*) (CYSTEINE CONJUGATE BETA-LYASE). NOV95 also has homology to the other proteins shown in the BLASTP data in Table 95D.

Table 95D. NOV95 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 12654031 gb AAH00819.1 AAH00819 (BC000819)	Similar to CG6950 gene product [<i>Homo sapiens</i>]	290	280/294 (95)	284/294 (96)	e-162
gi 5002565 emb CAB44334.1 (Y17462)	cysteine conjugate beta-lyase [<i>Takifugu rubripes</i>]	419	230/418 (55)	301/418 (71)	e-134
gi 4757928 ref NP_004050.1 (NM_004059)	cytoplasmic cysteine conjugate-beta lyase; glutamine-phenylpyruvate aminotransferase [<i>Homo sapiens</i>]	422	215/421 (51)	299/421 (70)	e-128
gi 15425868 gb AAK97625.1 AF395204_1 (AF395204)	kynurenine aminotransferase [<i>Aedes aegypti</i>]	477	224/420 (53)	284/420 (67)	e-126
gi 7299520 gb AAF54707.1 (AE003693)	CG6950 gene product [alt 3] [<i>Drosophila melanogaster</i>]	417	220/417 (52)	289/417 (68)	e-125

This BLASTP data is displayed graphically in the ClustalW in Table 95E. A multiple sequence alignment is given, with the NOV95 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 95D.

Table 95E. ClustalW Alignment of NOV95	
NOV95	(SEQ ID NO:300)
gi 12654031	(SEQ ID NO:774)
gi 5002565	(SEQ ID NO:775)
gi 4757928	(SEQ ID NO:776)
gi 15425868	(SEQ ID NO:777)
gi 7299520	(SEQ ID NO:778)
<div> <div>1020304050</div> <div>..... </div> </div>	
NOV95	-----
gi 12654031	-----
gi 5002565	-----
gi 4757928	-----
gi 15425868	MMFLRNHNSVGGAIRTAVVLQDLQFIVSNKSSALTGAUVSSVHRQQIRITMS
gi 7299520	-----
<div> <div>60708090100</div> <div>..... </div> </div>	
NOV95	-----MSLKFTNAKRIEGLDSNVWLEETKLAADPSVVNLGQGFDPDISPE
gi 12654031	-----MSRRIHANRTNGLDKNVWVEFTQLAAAYSKVNLGQGFDPDFAPP
gi 5002565	-----MAKQLQARRLDGIDYNPWVEFVKLASEHDVVNLGQGFDPDFAPP
gi 4757928	-----MAKQLQARRLDGIDYNPWVEFVKLASEHDVVNLGQGFDPDFAPP
gi 15425868	STSNETMHNKFDLPKRYOGSTKSVWVEFTQLAAAYKPLNLGQGFDPDYHAP
gi 7299520	-----MEKFDLPKRLOGSTPSVWVEFTQLAAAYKPLNLGQGFDPDAAPE
<div> <div>110120130140150</div> <div>..... </div> </div>	
NOV95	TYVKEELSKIAAIDS--LNQYTRGFGHPSSLVKALSYLYEKLYOKOIDSNK
gi 12654031	KEVQEAFCCHALNEGP--MMHQYTRAFGHVPLVKSIAKFFSRVIGHEIDPLE
gi 5002565	DFAVEAFQHAVSGDF--MLNQYTKTFGYPPLTKILASFFGELGQEIDPLR
gi 4757928	DFAVEAFQHAVSGDF--MLNQYTKTFGYPPLTKILASFFGELGQEIDPLR
gi 15425868	KYALNALAAAANSPPDPLANQYTRGFGHPRLVQALSPLYSLVDRTINPMT
gi 7299520	EMVTHSLADIAKEQNPLLEHQYTRGYGHVRLVNALSKLYSGLVGKELNPLS
<div> <div>160170180190200</div> <div>..... </div> </div>	
NOV95	EILVTVGAYGSLFNTIQALIDEG--QVILTVFFYDCYEPMVRMAGATPVFI
gi 12654031	-----MVRMAGATPVFI
gi 5002565	DILVTVGAYQALSAFQALTYEGDEVIIIEPFFDCYQPMVKMAGGQPVFI
gi 4757928	NVLTVGCGYQALSAFQALTYEGDEVIIIEPFFDCYQPMVKMAGGQPVFI
gi 15425868	EVLVTVGAYEALXATIQGHVDEGDEVIIIEPFFDCYEPMVKRAGGIPRFI
gi 7299520	DILITSQAYEALYSTIMGHVTVGDEVIIIEPFFDCYEPMVKMAGGQPVFI
<div> <div>210220230240250</div> <div>..... </div> </div>	
NOV95	PLRS--VSLGKRWSSSDWTLDPQELSKFNSKTKAILNTPHNPLGKVYN
gi 12654031	PLRSK--PVYGKRWSSSDWTLDPQELSKFNSKTKAILNTPHNPLGKVYN
gi 5002565	PLRPK--GDGSLVSSGDWVLSPEXLAGKFTPTKALVINTPNPLGKVYK
gi 4757928	SLKPGPIQNGELGSSSNWQDPEMELAGKFTSRTKALVINTPNPLGKVFS
gi 15425868	PLKP--NKTGGTISSADWVLDNNELEALFNEKTKMILNTPHNPLGKVMD
gi 7299520	PLKL--RKTEGPISSADWVLDNNELEALFNEKTKMILNTPHNPLGKVFT
<div> <div>260270280290300</div> <div>..... </div> </div>	

NOV95:	212	VYEWLVYSGNKHKLK	IATFPGMWERTITIGSAGKTF	SVTGKVGWSIG	-----PNHL	262
		Y	V+ G	IA	++ ++ S K F + G ++G + G	+ +
Sbjct:	132	AYAGGVFGGLDGASIAELLDEYDNL	LVVQSLSKNFGLAGKRLGGAAGGIVAGSAASFDRV			191
NOV95:	263	IKHLQTVQONTIYTCATPLQEALAQAF	WIDIKRMDDPECYFNSLPKELEV	KRDRMVRLL		322
		+ +	AT	A + + D	E + L +	
Sbjct:	192	SSQSRALLF-----ATSSAPPAVGAAI	VALILQD-----K	ERLERWLKELKK		233
NOV95:	323	SVGLKPIVPDGGYFIIADV	SIFIVLDPDLSDMKNN	EPYDYKFVKWMTKHQKLSAIPVSA		382
		+GL+ ++	G+ + D S I+ L	+ + +P S		
Sbjct:	234	MLGLRVLLSRAGFVLWLDPS	-GILPLWTFEDQA-----GLFSALLLEEHGVV	VPGE		284
NOV95:	383	FCNSETKSQFEK	FVRF	CFIKVSSL-LDAAEEI	IKA	416
		F	R	++ LD E I+A		
Sbjct:	285	FPTVPPGW-----GRISLAGL	TDETLD	ELLEAIRA		314

Aminotransferases share certain mechanistic features with other pyridoxal-phosphate dependent enzymes, such as the covalent binding of the pyridoxal-phosphate group to a lysine residue. On the basis of sequence similarity, these various enzymes can be grouped into subfamilies, one of which is the class I. Class I aminotransferases include cysteine conjugate beta-lyase. Living organisms employ a variety of metabolic pathways when detoxifying xenobiotic compounds, including the formation of cysteine S-conjugates via glutathione conjugation. Kidney cysteine conjugate beta-lyase (glutamine transaminase K, kyneurenine aminotransferase, EC 2.6.1.64) metabolises the cysteine conjugates of certain halogenated alkenes and alkanes to form reactive metabolites which can produce nephrotoxicity and neurotoxicity in experimental animals and man.

NOV95 is predicted to be expressed in at least the following tissues: kidney, liver, colon, gall bladder. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV95 is provided in Example 2.

The NOV95 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV95 nucleic acids encoding the aminotransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a cysteine conjugate beta-lyase-like protein includes the nucleic acid whose sequence is provided in Table 95A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 95A while still encoding a protein
5 that maintains its cysteine conjugate beta-lyase-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 95A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or
10 complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.
15 In the mutant or variant nucleic acids, and their complements, up to about 12% of the bases may be so changed.

The novel protein of the invention includes the cysteine conjugate beta-lyase-like protein whose sequence is provided in Table 95B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown
20 in Table 95B while still encoding a protein that maintains its cysteine conjugate beta-lyase-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 45% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic
25 methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV96

NOV96 includes three monocarboxylate transporter-like proteins, designated herein as
30 NOV96a, NOV96b and NOV96c.

NOV96a

The disclosed NOV96a (alternatively referred to herein as CG56918-01) includes the 1302 nucleotide sequence (SEQ ID NO:301) shown in Table 96A. A NOV96a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 9-11 and ends with a stop codon at nucleotides 1287-1289. The disclosed NOV96a maps to human chromosome 17.

Table 96A. NOV96a Nucleotide Sequence (SEQ ID NO:301)

```
CCGCTTAGATGGCGCGCAGGACAGAGCCCCGACGGGGGCTGGGGATGGGTGGTGGTGC
TCTCAGCGTTCTTCCAGTCGGCGCTTGTGTTTGGGGTGCTCCGCTCCTTTGGGGTCTTCT
TCGTGGAGTTTGTGGCGGCGTTGAGGAGCAGGCAGCGCGCTCTCCTGGATCGCCTCCA
TAGGAATCGCGGTGCAGCAGTTTGGGAGCCCGGTAGGCAGTGCCCTGAGCACGAAGTTCG
GGCCCAGGCCCCGTGGTGATGACTGGAGGCATCTTGGCTGCGCTGGGGATGCTGCTCGCT
CTTTTGCTACTTCTTTGACCCACCTATACCTGAGTATTGGGTGCTGTGACGGCTCTGGCT
GGGCTTTGACCTTCGCTCCGACCTGGCCTGCCTGCTGTTATTTTCTCGCCGACGAT
CCCTGGCCACCGGGCTGGCACTGACAGGCGTGGGCCTCTCTCCTTCACATTGCCCCCT
TTTTCAGTGGCTGCTCAGCCACTACGCCCTGGAGGGGGTCCCTGCTGCTGGTGTCTGCCC
TCTCCCTCCACCTAGTGGCCTGTGGTGTCTCTCCGCCCCACCCTCCCTGGCTGAGGACC
CTGCTGTGGGTGGTCCCAGGGCCCAACTCACCTCTCTCCTCCATCATGGCCCCCTCTCTCC
GTTACACTGTTGCCCTCACCTGATCAACACTGGCTACTTCATTCCCTACCTCCACCTGG
TGGCCCATCTCCAGGACCTGGATTGGGACCCACTACCTGCTGCCTTCTACTCTCAGTTG
TTGCTATTTCTGACCTCGTGGGGCGTGTGGTCTCCGGATGGCTGGGAGATGCAGTCCAG
GGCCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGGGTGTCACTAGCCCTGT
TCCCTGTAGCTCAGGCTCCACAGCCCTGGTGGCTCTGGCTGTGGCCTACGGCTTCACAT
CAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACCTAATAGGGACTAGAAGGA
TTTACTGTGGCCTGGGACTGTTGCAGATGGTAGAGAGCATCGGGGGGCTGTGGGGGCTC
CTCTCTCAGGCTACCTCCGGGATGTGACAGGCAACTACACGGCTTCTTTTGGTGGTGGCTG
GGGCTTCTCTCTTTTACGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTCTGCTCTCT
CAACTACTACCTCCGGGCCCCAGGACCTTGTAAACAGAAGCACTAGATACTAAAGTTCCCT
TACCCAAGGAGGGGCTGGAAGAGGACTGAACTCCACAGAGTC
```

A NOV96a polypeptide (SEQ ID NO:302) encoded by SEQ ID NO:301 is 426 amino acids in length and is presented using the one-letter amino acid code in Table 96B. The Psort profile for NOV96a predicts that this sequence is a Type IIIa membrane protein, has a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.6400. In alternative embodiments, a NOV96a polypeptide is located to the Golgi with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.3700. The Signal P predicts a likely cleavage site for a NOV96a peptide is between positions 34 and 35, *i.e.*, at the dash in the sequence VAA-FE.

Table 96B. NOV96a Polypeptide Sequence (SEQ ID NO:302)

```
MARRTEPPDGGWGVVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
AVQQFGSPVGSALSTKFGPRPVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
TFAPTLACLSCYFSRRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSAISL
HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
LQDLWDPLPAAFLLSVVAISDLVGRVVSGLGDAVPGPVTRLLMLWTLTGVSIALFPV
AQAPTALVALAVAYGFTSGALAPLAFSVLPELIGTRRIYCGLLQMVESIGLLGPPLS
GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFCSSTTSGPQDLVTEALDTKVPLPK
EGLEED
```


NOV96b

The disclosed NOV96b (alternatively referred to herein as CG56918-02) includes the 1294 nucleotide sequence (SEQ ID NO:303) shown in Table 96C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 1279-1281. The disclosed NOV96b maps to human chromosome 17.

Table 96C. NOV96b Nucleotide Sequence (SEQ ID NO:303)

ATGGCGCGCAGGACAGAGCCCCGACGGGGCTGGGGATGGGTGGTGGTCTCTCAGCG
TCTTCCAGTCGGCGCTTGTGTTTGGGGTGCTCCGCTCCTTTGGGGTCTTCTTCGTGGAG
TTTGTGGCGCGTGTGAGGAGCAGGCAGCGCGCTCTCCTGGATCGCCTCCATAGGAATC
GCGGTGCAGCAGTTTGGGAGCCCGGTAGGCAGTGCCCTGAGCACGAAGTTCGGGCCCCAGG
CCCGTGGTGATGACTGGAGGCATCTTGGCTGCGCTGGGGATGCTGCTCGCCTCTTTTGCT
ACTTCCCTGACCCACCTATACCTGAGTATTGGGTGCTGTGCTGAGGCTCTGGCTGGGCTTTG
ACCTTCGCTCCGACCTGGCCTGCCTGTCTGTTATTTTCTCGCCGACGATCCCTGGCC
ACCGGGCTGGCAGTACAGGCGTGGGCTCTCCTCCTTACATTGCCCCCTTTTCCAG
TGGCTGCTCAGCCACTACGCTGGAGGGGTCCCTGCTGCTGGTGTCTGCCCTCTCCCTC
CACCTAGTGGCCTGTGGTCTCTCCTCCGCCCACCTCCCTGGCTGAGGACCTGCTGTG
GGTGGTCCAGGGCCCAACTACCTCTCTCCTCCATCATGGCCCTTCTCCGTTACACT
GTTGCCCTCACCTGATCAACACTGGCTACTTTCATTCCTACCTCCACCTGGTGGCCCAT
CTCCAGGACCTGGATTGGGACCCACTACCTGCTGCCTTCTACTCTCAGTTGTTGCTATT
TCTGACCTCGTGGGCGTGTGGTCTCCGGATGGCTGGGAGATGAGTCCCAGGGCCTGTG
ACACGACTCCTGATGCTCTGGACCACTTGACTGGGGTGCTACTAGCCCTGTTCCCTGTA
GCTCAGGCTCCCACAGCCCTGGTGGCTCTGGCTGTGGCTACGGCTTACATCAGGGGCT
CTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACATAAGGGACTAGAAGGATTTACTGT
GGCCTGGGACTGTTGAGATGGTAGAGAGCATCGGGGGGCTGCTGGGGCCTCTCTCTCA
GGCTACCTCCGGGATGTGACAGGCAACTACACGGCTTCTTTGTGGTGGCTGGGGCCTTC
CTTCTTTCAGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTTCTGCTCCTCAACTACT
ACCTCCGGGCCCCAGGACCTTGTAACAGAAGCACTAGATACTAAAGTTCCCTACCCAAG
GAGGGACTGGAAGAGGACTGAACTCCACAGAGTC

A NOV96b polypeptide (SEQ ID NO:304) encoded by SEQ ID NO:303 is 426 amino acids in length and is presented using the one-letter amino acid code in Table 96D. The Psort profile for NOV96b predicts that this sequence is likely to be a Type IIIa membrane protein, has a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.6400. In alternative embodiments, a NOV96b polypeptide is located to the Golgi with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.3700. The Signal P predicts a likely cleavage site for a NOV96b peptide is between positions 44 and 45, *i.e.*, at the dash in the sequence VAA-FE.

Table 96D. NOV96b Polypeptide Sequence (SEQ ID NO:304)

MARRTEPPDGGWGVVVLSAFFQSALVFGVLRSGVFVEFVAAFEQAAARVSWIASIGI
AVQQFGSPVGSALSTKFGPRPVVMTGGILAALGMLLASFATSLTHLYLSIGLLSGSGWAL
TFAPTLACLSCYFSRRRLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVLSLSL
HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVHA
LQDLWDPLPAAFLSVVAISDLVGRVVSGLGDAVPGPVTRLLMLWTLTGVSALFPV
AQAPTALVALAVAYGFTSGALAPLAFSVLPeligTRRIYCGLGLLQMVESIGLLGPPLS
GYLRDVTGNYTASFVVAGAFLLSGSGILLTLPHFCSSTTSGPQDLVTEALDTRVPLPK

NOV96c

The disclosed NOV96c (alternatively referred to herein as CG56918-03) includes the 1445 nucleotide sequence (SEQ ID NO:305) shown in Table 96E. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 9-11 and ends with a stop codon at nucleotides 1287-1289. The disclosed NOV96c maps to human chromosome 17.

Table 96E. NOV96c Nucleotide Sequence (SEQ ID NO:305)

```
CCGCTTAGATGGCGCGCAGGACAGAGCCCCGACGGGGGCTGGGGATGGGTGGTGGTGC
TCTCAGCGTTCTTCCAGTCGGCGCTTGTGTTTGGGGTGCTCCGCTCCTTTGGGGTCTTCT
TCGTGGAGTTTGTGGCGCGTTTGGAGAGCAGGCAGCGCGCTCTCTGGATCGCCTCCA
TAGGAATCGCGGTGCAGCAGTTTGGGAGCCCGGTAGGCAGTGCCCTGAGCACGAAGTTCG
GGCCAGGCCCGTGGTGATGACTGGAGGCATCTTGGCTGCGCTGGGGATGCTGCTCGCCT
CTTTTGCTACTTCTTGGACCCACCTATACCTGAGTATTGGGTGCTGTCAGGCTCTGGCT
GGGCTTTGACCTTCGCTCCGAGCCTGGCCTGCCTGCTCCTGTTATTCTCTCGCCGACGAT
CCCTGGCCACCGGGCTGGCACTGACAGGCGTGGGCCTCTCCTCCTTACATTTGCCCCCT
TTTCCAGTGGCTGCTCAGCCACTACGCCTGGAGGGGGTCCCTGCTGCTGGTGTCTGCC
TCTCCCTCCACCTAGTGGCCTGTGGTGCTCTCCTCCGCCACCCTCCCTGGCTGAGGACC
CTGCTGTGGGTGGTCCCGAGGGCCCAACTCACCTCTCTCCTCCATCATGGCCCCCTTCTCC
GTTACACTGTTGCCCTCACCTGATCAACACTGGCTACTTCATTCCCTACCTCCACCTGG
TGGCCCATCTCCAGGACCTGGATTGGGACCCACTACCTGCTGCCTTCTACTCTCAGTTG
TTGCTATTCTGACCTCGTGGGGCTGTGGTCTCCGGATGGCTGGGAGATGCAGTCCCAG
GGCTGTGACACGACTCCTGATGCTCTGGACCACCTTGACTGGGGTGTCACTAGCCCTGT
TCCCTGTAGCTCAGGCTCCACAGCCCTGGTGGCTCTGGCTGTGGCTACGGCTTCACAT
CAGGGGCTCTGGCCCCACTGGCCTTCTCCGTGCTGCCTGAACTAATAGGGACTAGAAGGA
TTTACTGTGGCTGGGACTGTTGCAGATGATAGAGAGCATCGGGGGGCTGCTGGGGCCTC
CTCTCTCAGGCTACCTCCGGGATGTGTGAGGCAACTACACGGCTTCTTTGTGGTGGCTG
GGGCCTTCTTCTTTCAGGGAGTGGCATTCTCCTCACCTGCCCCACTTCTTCTGCTTCT
CACTACTACCTCCGGGCCTCAGGACCTTGTAACAGAAGCACTAGATACTAAAGTTCCTCC
TACCCAAGGAGGGGCTGGAAGAGGACTGAACTCCACAGAGTCAGGCCAGAAAGCCAAAG
CTTGACAGCTCCAGGTCTTCTCTTGCCACGTCTTGGTCTCCACAGAACACAGTGCCTTA
AGATTCTTGATCTGCCTCCCCCTAGAGCAGGCTGGGGCTCCTGCAATGTGTGTGCCAAC
CCTTT
```

The NOV96c polypeptide (SEQ ID NO:306) encoded by SEQ ID NO:305 is 426 amino acids in length and is presented using the one-letter amino acid code in Table 96F. The Psort profile for NOV96c predicts that this sequence is a Type III a membrane protein, has a signal peptide, and is likely to be localized to the plasma membrane with a certainty of 0.6400. In alternative embodiments, a NOV96c polypeptide is located to the Golgi with a certainty of 0.4600, or to the endoplasmic reticulum (membrane) with a certainty of 0.3700. The Signal P predicts a likely cleavage site for a NOV96c peptide is between positions 34 and 35, *i.e.*, at the dash in the sequence VAA-FE.

Table 96F. NOV96c Polypeptide Sequence (SEQ ID NO:306)

MARRTEPPDGGWGVVLSAFFQSALVFGVLRSGVFFVEFVAAFEEQAARVSWIASIGI
AVQQFGSPVGSALSTKFGPRPVMTGGILAAAGMLLASFATSLTHLYLSIGLLSGSGWAL
TFAPSLACLSYFSRRSLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS
HLVACGALLRPPSLAEDPAVGGPRAQLTSLHHGPFRLRYTVALTLINTGYFIPYLHLVAH
LQDLWDPLPAAFLLSVVAISDLVGRVVSGLGDAVPGPVTRLLMLWTTLTGVSALALFPV
AQAPTALVALAVAYGFTSGALAPLAFSVLPELIGTRRIYCGLGLLQMIESIGLLGPPLS
GYLRDVSGNYTASFVVAGAFLLSGSGILLTLPHFFCFSTTTSGPQDLVTEALDTKVPLPK
EGLEED

A BLAST analysis of NOV96 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV96a had high homology to other proteins as shown in Table 96G.

5

Table 96G. BLASTX results from PatP database for NOV96a

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AAU01618 Human secreted protein	940	3.0e-94
patp:AAM93737 Human polypeptide	940	3.0e-94
patp:AAB88570 Human hydrophobic domain containing protein	620	2.5e-60
patp:AA31642 Human transport-associated protein-4	602	2.0e-58
patp:AAU01586 Human secreted protein related to gene #26	357	1.8e-32

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 420 of 702 bases (59%) identical to a gb:GENBANK-ID:RNU87627|acc:U87627.1 mRNA from *Rattus norvegicus* (*Rattus norvegicus* putative monocarboxylate transporter (MCT3) mRNA). The full amino acid sequence of the protein of the invention was found to have 89 of 191 amino acid residues (46%) identical to, and 119 of 191 amino acid residues (62%) similar to, the 504 amino acid residue ptmr:SPTREMBL-ACC:O95907 protein from *Homo sapiens* (Human) (DJ1039K5.2 (SIMILAR TO MONOCARBOXYLATE TRANSPORTER (MCT3)). NOV96 also has homology to the other proteins shown in the BLASTP data in Table 96H.

15

Table 96H. NOV96 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17491104 ref XP_064368.1 (XM_064368)	similar to solute carrier family 16 (monocarboxylic acid transporters), member 8 (H. sapiens) [<i>Homo sapiens</i>]	427	424/427 (99)	425/427 (99)	0.0

gi 14754801 ref XP_044738.1 (XM_044738)	solute carrier family 16 (monocarboxylic acid transporters), member 7 [Homo sapiens]	478	141/410 (34)	213/410 (51)	2e-57
gi 4759120 ref NP_004722.1 (NM_004731)	solute carrier family 16 (monocarboxylic acid transporters), member 7; monocarboxylate transporter 2 [Homo sapiens]	478	141/410 (34)	213/410 (51)	2e-57
gi 2497855 sp Q63344 MOT2_RA T	MONOCARBOXYLATE TRANSPORTER 2 (MCT 2)	489	140/410 (34)	215/410 (52)	7e-57
gi 1432167 gb AAB04023.1 (U62316)	monocarboxylate transporter 2 [Rattus norvegicus]	489	141/410 (34)	215/410 (52)	8e-57

This BLASTP data is displayed graphically in the ClustalW in Table 96I. A multiple sequence alignment is given, with the NOV96a, b, and c proteins being shown on lines 1, 2, and 3 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 96H.

Table 96I. ClustalW Alignment of NOV96

NOV96a	(SEQ ID NO:302)
NOV96b	(SEQ ID NO:304)
NOV96c	(SEQ ID NO:306)
gi 17491104	(SEQ ID NO:779)
gi 14754801	(SEQ ID NO:780)
gi 4759120	(SEQ ID NO:781)
gi 2497855	(SEQ ID NO:782)
gi 1432167	(SEQ ID NO:783)
	10 20 30 40 50
NOV96a	-----MARRTEPPDGGGWG-VVVLSAFFQSALVEGVLRSFGVF
NOV96b	-----MARRTEPPDGGGWG-VVVLSAFFQSALVEGVLRSFGVF
NOV96c	-----MARRTEPPDGGGWG-VVVLSAFFQSALVEGVLRSFGVF
gi 17491104	-----MARRTEPPDGGGWG-VVVLSAFFQSALVEGVLRSFGVF
gi 14754801	MPP-----MPSAPPVHPPDGGGWG-IVVGAFTLSIGFSYAFPKAVTVF
gi 4759120	MPP-----MPSAPPVHPPDGGGWG-IVVGAFTLSIGFSYAFPKAVTVF
gi 2497855	MPSESVKATAAPPFPPLPPDGGGWG-VVVCASFISIGFSYAFPKAVTVF
gi 1432167	MPSESVKATAAPPFPPLPPDGGGWG-VVVCASFISIGFSYAFPKAVTVF
	60 70 80 90 100
NOV96a	FVEFVAARFEEQAARVSWIASIGTAVQQFCSPVGSALSTKFGPRPVVMTGG
NOV96b	FVEFVAARFEEQAARVSWIASIGTAVQQFCSPVGSALSTKFGPRPVVMTGG
NOV96c	FVEFVAARFEEQAARVSWIASIGTAVQQFCSPVGSALSTKFGPRPVVMTGG
gi 17491104	FVEFVAARFEEQAARVSWIASIGTAVQQFCSPVGSALSTKFGPRPVVMTGG
gi 14754801	RKEIQQIHHTTYSETAWISSIMLAVMYACGPVSSVLVNNKYGSRPVVVIAGG
gi 4759120	RKEIQQIHHTTYSETAWISSIMLAVMYACGPVSSVLVNNKYGSRPVVVIAGG
gi 2497855	RNDIKDIEKTTSSQIAWISSIMLAVMYACGPVSSVLVNNKYGSRPVVVIAGG
gi 1432167	RNDIKDIEKTTSSQIAWISSIMLAVMYACGPVSSVLVNNKYGSRPVVVIAGG
	110 120 130 140 150
NOV96a	ITLAALGMILLASFATSIITHLYLSIGLLSGSGWALTFAFTLACLSYFSRRR
NOV96b	ITLAALGMILLASFATSIITHLYLSIGLLSGSGWALTFAFTLACLSYFSRRR
NOV96c	ITLAALGMILLASFATSIITHLYLSIGLLSGSGWALTFAFTLACLSYFSRRR
gi 17491104	ITLAALGMILLASFATSIITHLYLSIGLLSGSGWALTFAFTLACLSYFSRRR

gi 14754801	LLCCLGMVLASFSSSVVQLYLTMGFTTGLGLAFNLOPALTITIGKYFYRKR
gi 4759120	LLCCLGMVLASFSSSVVQLYLTMGFTTGLGLAFNLOPALTITIGKYFYRKR
gi 2497855	LLCCTGMILASFSSSVIELYLTVCFTGGLGLAFNLOPALTITIGKYFYRKR
gi 1432167	LLCCTGMILASFSSSVIELYLTVCFTGGLGLAFNLOPALTITIGKYFYRKR

	160	170	180	190	200
NOV96a	SLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS	SLHLVACGA			
NOV96b	SLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS	SLHLVACGA			
NOV96c	SLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS	SLHLVACGA			
gi 17491104	SLATGLALTGVGLSSFTFAPFFQWLLSHYAWRGSLLLVSALS	SLHLVACGA			
gi 14754801	PMANGLAMAGSPVFLSSLAPFNOYLFNTFGWKGSFLILGSL	LLNACVAGS			
gi 4759120	PMANGLAMAGSPVFLSSLAPFNOYLFNTFGWKGSFLILGSL	LLNACVAGS			
gi 2497855	PLANGFAMAGSPVFLSLAPFNOYLFNSYGWKGSFLILGAI	FLHSCVACC			
gi 1432167	PLANGFAMAGSPVFLSLAPFNOYLFNSYGWKGSFLILGAI	FLHSCVACC			

	210	220	230	240	250
NOV96a	LLRPPSLAEDPAVGGPRAQLTS				LLH
NOV96b	LLRPPSLAEDPAVGGPRAQLTS				LLH
NOV96c	LLRPPSLAEDPAVGGPRAQLTS				LLH
gi 17491104	LLRPPSLAEDPAVGGPRAQLTS				LLH
gi 14754801	LMRPLGPNQTTSKSKNKTGKTEDDSSPKKI	IKTKKSTWEKVNKYLD			FSFK
gi 4759120	LMRPLGPNQTTSKSKNKTGKTEDDSSPKKI	IKTKKSTWEKVNKYLD			FSFK
gi 2497855	LMRPVGPSPRAAKSKSKVGSRDSS	TKR--LSKVSTA			ELNRFDFGLFT
gi 1432167	LMRPVGPSPRAAKSKSKVGSRDSS	TKR--LSKVSTA			ELNRFDFGLFT

	260	270	280	290	300
NOV96a	HGPFRLRYTVALTLLNTGYFTIPYLHLVAHLQDL	DWDLPLPA	AFLLSVVA		ISD
NOV96b	HGPFRLRYTVALTLLNTGYFTIPYLHLVAHLQDL	DWDLPLPA	AFLLSVVA		ISD
NOV96c	HGPFRLRYTVALTLLNTGYFTIPYLHLVAHLQDL	DWDLPLPA	AFLLSVVA		ISD
gi 17491104	HGPFRLRYTVALTLLNTGYFTIPYLHLVAHLQDL	DWDLPLPA	AFLLSVVA		ISD
gi 14754801	HRGFLLIYLSGNVIMFLGFAPIIFLAPYAKDQ	GIDEYS	AFLLSVMA		FVD
gi 4759120	HRGFLLIYLSGNVIMFLGFAPIIFLAPYAKDQ	GIDEYS	AFLLSVMA		FVD
gi 2497855	HRGFLLIYLSGNVIMFLGFAPIIFLAPYAKD	KGVDDYNS	AFLLSVMA		FID
gi 1432167	HRGFLLIYLSGNVIMFLGFAPIIFLAPYAKD	KGVDDYNS	AFLLSVMA		FID

	310	320	330	340	350
NOV96a	LVGRVVSGLWG--DAVPGPVTRLLMLWTTLTG	VSALFPV	ACAPTAL		VAL
NOV96b	LVGRVVSGLWG--DAVPGPVTRLLMLWTTLTG	VSALFPV	ACAPTAL		VAL
NOV96c	LVGRVVSGLWG--DAVPGPVTRLLMLWTTLTG	VSALFPV	ACAPTAL		VAL
gi 17491104	LVGRVVSGLWG--DAVPGPVTRLLMLWTTLTG	VSALFPV	ACAPTAL		VAL
gi 14754801	MFARPSVGLTANTSKYIRPRIQYFFSF	AIMFNGV	CHLLCPL		ADYTSLVLY
gi 4759120	MFARPSVGLTANTSKYIRPRIQYFFSF	AIMFNGV	CHLLCPL		ADYTSLVLY
gi 2497855	MFARPSVGLTANTSLIRPRIQYLF	SVAIMFTG	ICHLLCPL		AHSYTALVVY
gi 1432167	MFARPSVGLTANTSLIRPRIQYLF	SVAIMFTG	ICHLLCPL		AHSYTALVVY

	360	370	380	390	400
NOV96a	AVAYGFTSGALAPLAFSVLP	ELIGTRRIYCGLGL	QMVESIG	GGLLGP	PLS
NOV96b	AVAYGFTSGALAPLAFSVLP	ELIGTRRIYCGLGL	QMVESIG	GGLLGP	PLS
NOV96c	AVAYGFTSGALAPLAFSVLP	ELIGTRRIYCGLGL	QMVESIG	GGLLGP	PLS
gi 17491104	AVAYGFTSGALAPLAFSVLP	ELIGTRRIYCGLGL	QMVESIG	GGLLGP	PLS
gi 14754801	AVFFGLGFGSVSSVLFETLMDLV	GAPRFSSAVGLVT	IVECCPV	LLGP	PLA
gi 4759120	AVFFGLGFGSVSSVLFETLMDLV	GAPRFSSAVGLVT	IVECCPV	LLGP	PLA
gi 2497855	VIFFGIGFGSISLLFECLMDQV	GASRFSSAVGLVT	IVECCPV	LLGP	PLA
gi 1432167	VIFFGIGFGSISLLFECLMDQV	GASRFSSAVGLVT	IVECCPV	LLGP	PLA

	410	420	430	440	450
NOV96a	CYL RDVTCNYTASFVAGAF	LLSGSGILL	TLPHFFCS	STTTSG	PD
NOV96b	CYL RDVTCNYTASFVAGAF	LLSGSGILL	TLPHFFCS	STTTSG	PD
NOV96c	CYL RDVTCNYTASFVAGAF	LLSGSGILL	TLPHFFCF	STTTSG	PD
gi 17491104	CYL RDVTCNYTASFVAGAF	LLSGSGILL	TLPHFFCF	STTTSG	PD
gi 14754801	GKLVDTGEYKYMMSCGAI	VVAASVWLL	IGNAINYR	LLAKER	KREENARQ
gi 4759120	GKLVDTGEYKYMMSCGAI	VVAASVWLL	IGNAINYR	LLAKER	KREENARQ
gi 2497855	GKLLDITGOYKYLNTASGI	VVLSGGIYLL	ICNAINYR	LLLEKER	KREKARR

gi 1432167	CKLLDITGOMKYLVIASGIVVLSSGIYLLICNAINYRLLEKERKREKARR
	460 470 480 490
NOV96a	A-----LDTKVPLPKEGLEED-----
NOV96b	A-----LDTKVPLPKEGLEED-----
NOV96c	A-----LDTKVPLPKEGLEED-----
gi 17491104	A-----LDTKVPLPKEGLEED-----
gi 14754801	K-----TRESEPLSKSKHSEDEVNVKVSNAQSVTSER--ETNI
gi 4759120	K-----SRESEPLSKSKHSEDEVNVKVSNAQSVTSER--ETNI
gi 2497855	KKSASQASKEMEALSRKQ-DDVTVKVSNTHNPPSDRDKESSI
gi 1432167	KKSASQASKEMEALSRKQ-DDVTVKVSNTHNPPSDRDKESSI

Monocarboxylates such as lactate and pyruvate play a central role in cellular metabolism and metabolic communication between tissues. Essential to these roles is their rapid transport across the plasma membrane, which is catalysed by a recently identified family of proton-linked monocarboxylate transporters (MCTs). Nine MCT-related sequences have so far been identified in mammals, each having a different tissue distribution, whereas six related proteins can be recognized in *Caenorhabditis elegans* and four in *Saccharomyces cerevisiae*.

Direct demonstration of proton-linked lactate and pyruvate transport has been demonstrated for mammalian MCT1-MCT4, but only for MCT1 and MCT2 have detailed analyses of substrate and inhibitor kinetics been described following heterologous expression in *Xenopus* oocytes. MCT1 is ubiquitously expressed, but is especially prominent in heart and red muscle, where it is up-regulated in response to increased work, suggesting a special role in lactic acid oxidation. By contrast, MCT4 is most evident in white muscle and other cells with a high glycolytic rate, such as tumour cells and white blood cells, suggesting it is expressed where lactic acid efflux predominates. MCT2 has a ten-fold higher affinity for substrates than MCT1 and MCT4 and is found in cells where rapid uptake at low substrate concentrations may be required, including the proximal kidney tubules, neurons and sperm tails. MCT3 is uniquely expressed in the retinal pigment epithelium. MCT1 and MCT4 have been shown to interact specifically with OX-47 (CD147), a member of the immunoglobulin superfamily with a single transmembrane helix. This interaction appears to assist MCT expression at the cell surface (Halestrap and Price, 1999, *Biochem. J.* vol.343: 281-99).

NOV96 is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, foreskin, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources,

public EST sources, literature sources, and/or RACE sources. Further expression data for NOV96 is provided in Example 2.

The NOV96 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV96 nucleic acids encoding the monocarboxylate transporter-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a monocarboxylate transporter-like protein includes the nucleic acid whose sequence is provided in Table 96A, 105C, or 105E, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 96A, 105C, or 105E while still encoding a protein that maintains its monocarboxylate transporter-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 96A, 105C, or 105E, including nucleic acid fragments that are complementary to any of the nucleic acids just described. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 41% of the bases may be so changed.

The novel protein of the invention includes the monocarboxylate transporter-like protein whose sequence is provided in Table 96B, 105D, or 105F. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 96B, 105D, or 105F while still encoding a protein that maintains its monocarboxylate transporter-like activities and physiological functions, or a functional

fragment thereof. In the mutant or variant protein, up to about 54% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV97

NOV97 includes six carboxypeptidase-like proteins, designated herein as NOV97a, NOV97b, NOV97c, NOV97d, NOV97e, and NOV97f.

NOV97a

The disclosed NOV97a (alternatively referred to herein as CG57070-01) includes the 1279 nucleotide sequence (SEQ ID NO:307) shown in Table 97A. A NOV97a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 1258-1260. The disclosed NOV97a maps to human chromosome 7q 31.

Table 97A. NOV97a Nucleotide Sequence (SEQ ID NO:307)

```
ATGCGGGGGTTGCTGGTGTGAGTGCTCTGTTGGGGGCTGTCTTTGGCAAGGAGGACTTT
GTGGGGCATCAGGTGCTCCGAATCTCTGTAGCCGATGAGGCCAGGTACAGAAGGTGAAG
GAGCTGGAGGACCTGGAGCACCTGCAGCTGGACTTCTGGCGGGGGCTGCCACCCTGGC
TCCCCCATCGACGTCCGAGTGCCCTTCCCCAGCATCCAGGCGGTCAAGATCTTTCTGGAG
TCCCACGGCATCAGCTATGAGACCATGATCGAGGACGTGCAGTCGCTGCTGGACGAGGAG
CAGGAGCAGATGTTTCGCTTCCGGTCCCCGGGCGCGCTCCACCGACACTTTTAACTACGCC
ACCTACCACACCCTGGAGGAGGTGTATAGCTGGATTGACAACCTTTGTAATGGAGCATCC
GATATTGTCTCAAAAATTCAGATTGGCAACAGCTTTGAAAACAGTCCATTCTTGTCCTG
AAGTTCAGCACTGGAGGTCTCGGCACCCAGCCATCTGGATCGACACTGGAATCACTCC
CGGAGTGGATCACCATGCCACCGGCATCTGGACTGCCAATAAGATTGTCACTGATTAT
GGCAAAGACCGTGTCTGACAGACATACTGAATGCCATGGACATCTTCATAGAGCTCGTC
ACAAACCCTGATGGGTTTGCTTTTACCCACAGCATGAACCGCTTATGGCGGAAGAACAAG
TCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGGAAGTGAAGTCGGGT
TTTGGAGGAATGGTTCTAACAGCAACCCCTGCTCAGAACTTATCACGGGCCCTCCCT
CAGTCGGAGCCGGAGGTGGCTGCCATAGTGAACCTTCATCACAGCCCATGGCAACTTCAAG
GCTCTGATCTCCATCCACAGCTACTCTCAGATGCTTATGTACCCTTACGGCCGATTGCTG
GAGCCCGTTTCAAATCAGAGGGAGTTGTACGATCTTGCCAAGGATGCGGTGGAGGCCTTG
TATAAGGTCCATGGGATCGAGTACATTTTGGCAGCATCAGCACCACCCTCTATGTGGCC
AGTGGGATCACCGTCGACTGGGCCCTATGACAGTGGCATCAAGTACGCCTTCAGCTTTGAG
CTCCGGGACACTGGGCAGTATGGCTTCTGCTGCCGGCCACACAGATCATCCCCACGGCC
CAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACACCTGAATCACCCCTACTAG
CAGCACGACTGAGGGCAGG
```

A NOV97a polypeptide (SEQ ID NO:308) encoded by SEQ ID NO:307 is 419 amino acids in length and is presented using the one-letter amino acid code in Table 97B. The Psort profile for NOV97a predicts that this sequence has a signal peptide and is likely to be

localized outside the cell with a certainty of 0.3703. In alternative embodiments, a NOV97a polypeptide is located to lysosomes with a certainty of 0.46200. The Signal P predicts a likely cleavage site for a NOV97a peptide is between positions 16 and 17, *i.e.*, at the dash in the sequence VFG-KE.

5

Table 97B. NOV97a Polypeptide Sequence (SEQ ID NO:308)
MRGLLVLSVLLGAVFGKEDFVGHQVLRISVADEAQVKVKELEDLEHLQLDFWRGPAHPG SPIDVRVPFPSIQAVKIFLESHGISYETMIEDVQSLLDDEEQQMFAFRSRARSTDTFNYA TYHTLEEVYSWIDNFMVMEHSDIVSKIQIGNSFENQSILVLKFPSTGGSRHPAIWIDTGIHS REWITHATGIWTANKIVSDYGKDRVLTDLNAMDIFIELVTNPDGFAFTHSMNRLWRKNK SIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSEYHGPSPQSEPEVAAIVNFITAHGNFK ALISIHYSQMLMYPYGRLLPEVSNQRELYDLAKDAVEALYKVHGIEYIFGSIISTTLYVA SGITVDWAYDSGIKYAFSFEELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY

NOV97b

The disclosed NOV97b (alternatively referred to herein as CG57070-02) includes the 1291 nucleotide sequence (SEQ ID NO:309) shown in Table 97C. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a stop codon at nucleotides 1270-1272. The disclosed NOV97b maps to human chromosome 7q31.

10

Table 97C. NOV97b Nucleotide Sequence (SEQ ID NO:309)
ATGCGGGGGTTGCTGGTGTGAGTGTCTGTTGGGGGCTGTCTTTGGCAAGGAGGACTTT GTGGGGCATCAGGTGCTCCGAATCTCTGTAGCCGATGAGGCCAGGTACAGAAGGTGAAG GAGCTGGAGGACCTGGAGCACCTGCAGGTGGACTTCTGGCGTGGCCAGCCAGGCCAGC CTCCCTGTGGATATGAGAGTTCCTTCTCTGAAGTGAAGACATCAAAGCTTATCTGGAG TCTCATGGACTTGCTTACAGCATCATGATAAAGGACATCCAGGTGAAGCCCCAGGTGCTG CTGGATGAGGAAAGACAGGCCATGGCGAAATCCCGCCGGCTGGAGCGCAGCACCAACAGC TTCAGTTACTCATACACACCTGGAGGAGGTATATAGCTGGATTGACAACTTTGTA ATGGAGCATTCGATATTGTCTCAAAAATTCAGATTGGCAACAGCTTTGAAAACAGTCC ATTCTTGTCTGAAGTTCAGCACTGGAGGTTCTCGGCACCCAGCCATCTGGATGCACT GGAATTCACCTCCCGGAGTGGATCACCCATGCCACCGGCATCTGGACTGCCAATAAGATT GTCAGTGATTATGGCAAAGACCGTGTCTGACAGACATACTGAATGCCATGGACATCTTC ATAGAGCTCGTCACAAACCCTGATGGGTTTGCTTTTACCCACAGCATGAACCGCTTATGG CGGAAGAACAAGTCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGGAAC TGGAAGTCGGGTTTGGAGGAAATGGTCTTAACAGCAACCCTGCTCAGAAACTTATCAC GGGCCCTCCCTCAGTCGGAGCCGAGGTGGCTGCCATAGTGAACCTCATCACAGCCCAT GGCAACTTCAAGGCTCTGATCTCCATCCACAGCTACTCTCAGATGCTTATGTACCTTAC GGCCGATTGCTGGAGCCCCGTTTCAAATCAGAGGGAGTTGTACGATCTTGCCAAGGATGCG GTGGAGGCCCTTGTATAAGGTCCATGGGATCGAGTACATTTTGGCAGCATCAGCACCAAC CTTATGTGGCCAGTGGGATCACCGTCCGACTGGGCTATGACAGTGGCATCAAGTACGCC TTCAGCTTTGAGCTCCGGGACACTGGGCAGTATGGCTTCTGCTGCCGCCACACAGATC ATCCCCACGGCCAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACACCTGAAT CACCCCTACTAGCAGCACGACTGAGGCGAG

A NOV97b polypeptide (SEQ ID NO:310) encoded by SEQ ID NO:309 is 423 amino acids in length and is presented using the one-letter amino acid code in Table 97D. The Psort profile for NOV97b predicts that this sequence has a signal peptide and is likely to be

15

localized outside the cell with a certainty of 0.3703. In alternative embodiments, a NOV97b polypeptide is located to lysosomes with a certainty of 0.1900. The Signal P predicts a likely cleavage site for a NOV97b peptide is between positions 16 and 17, *i.e.*, at the dash in the sequence VFG-KE.

5

Table 97D. NOV97b Polypeptide Sequence (SEQ ID NO:310)

MRGLLVLSVLLGAVFGKEDFVGHQVLRISVADEAQVQKVELEDLEHLQVDFWRGPARPS
LPVDMRVPFSELKDIKAYLESHGLAYSIMIKDIQVKPQVLLDEERQAMAKSRRLERSTNS
FSYSSYHTLEEVYSWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDT
GIHSREWTHATG IWTANKIVSDYGKDRVLT DILNAMDIFIELVTNPDGFAFTHSMNRLW
RKNKSIRPGIFCIGVDLNRNWKSGFGNGSNSNPCSETYHGPSPQSEPEVAIVNFITAH
GNFKALISIHYSQMLMYPYGRLLPEVSNQRELYDLAKDAVEALYKVHGIEYIFGSISTT
LYVASGITVDWAYDSGIKYAFSFE LRDTGQYGFLLPATQIIPTAQETWMLRTIMEHTLN
HPY

NOV97c

The disclosed NOV97c (alternatively referred to herein as CG57070-03) includes the 1344 nucleotide sequence (SEQ ID NO:311) shown in Table 97E. A NOV97a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 25-27 and ends with a stop codon at nucleotides 1327-1329.

10

Table 97E. NOV97c Nucleotide Sequence (SEQ ID NO:311)

TGAAGCTCACCAGGAGGAAGAAGCATGCAGGGCACCCCTGGAGGCGGGACGCGCCCTGGG
CCATCCCCCGTGGACAGGCGGACACTCCTGGTCTTCAGCTTTATCCTGGCAGCAGCTTTG
GGCCAAATGAATTTACAGGGCAGGTTCTTCGAGTCTGGCCAAAGATGAGAAGCAGCTT
TCACTTCTCGGGGATCTGGAGGGCTGAAACCCAGAAAGGTGGACTTCTGGCGTGGCCCA
GCCAGGCCCAGCCTCCCTGTGGATATGAGAGTTCCTTTCTGAACTGAAAGACATCAAA
GCTTATCTGGAGTCTCATGGACTTGCTTACAGCATCATGATAAAGGACATCCAGGTGCTG
CTGGATGAGGAAAGACAGGCCATGGCGAAATCCCGCCGGCTGGAGCGCAGCACCAACAGC
TTCAGTTACTCATCATACACACCCTGGAGGAGATATAGCTGGATTGACAACTTTGTA
ATGGAGCATTCCGATATTGTCTCAAAAATTCAGATTGGCAACAGCTTTGAAAACAGTCC
ATCTTGTCTGAAGTTCAGCACTGGAGGTTCTCGGCACCCAGCCATCTGGATTGACACT
GGAATTCACCTCCCGGAGTGGATCACCCATGCCACCGGCATCTGGACTGCCAATAAGATT
GTCAGTGATTATGGCAAAGACCGTGTCTGACAGACATACTGAATGCCATGGACATCTTC
ATAGAGCTCGTCACAAACCTGATGGGTTTGCTTTTACCCACAGCATGAACCGCTTATGG
CGGAAGAACAAGTCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGGAAC
TGGAAGTCGGGTTTGGAAATGGTTCTAACAGCAACCCCTGCTCAGAACTTATCACGGG
CCCTCCCCCTCAGTCGGAGCCGAGGTGGCTGCCATAGTGAACCTCATCACAGCCCATGGC
AACTTCAAGGCTCTGATCTCCATCCACAGCTACTCTCAGATGCTTATGTACCCTTACGGC
CGATTGTCTGGAGCCCGTTTCAAAATCAGAGGGAGTTGTACGATCTTGCCAAGGATGCGGTG
GAGGCCTTGATAAGGTCCATGGATCGAGTACATTTTGGCAGCATCAGCACCAACCTC
GATGTGGCCAGTGGGATCACCGTGGAGTGGGCTATGACAGTGGCATCAAGTACGCCTTC
AGCTTTGAGTCCGGGACACTGGGCAGTATGGCTTCTGCTGCCGGCCACACAGATCATC
CCCACGGCCCAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACACCCTGAATCAC
CCCTACTAGCAGCAGCTGAGGG

A NOV97c polypeptide (SEQ ID NO:312) encoded by SEQ ID NO:311 is 434 amino acids in length and is presented using the one-letter amino acid code in Table 97F. The Psort

15

profile for NOV97c predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.5851. In alternative embodiments, a NOV97c polypeptide is located to lysosomes with a certainty of 0.4366. The Signal P predicts a likely cleavage site for a NOV97c peptide is between positions 33 and 34, *i.e.*, at the dash in the sequence ALG-QM.

Table 97F. NOV97c Polypeptide Sequence (SEQ ID NO:312)

MQGTPGGGTRPGPSPVDRRTLLVFSFILAAALGQMNFTGQVLRVLAKDEKQLSLGDLLEG
LKPQKVDWFRGPARPSLPVDMRVFPFSELKDIKAYLESHGLAYSIMIKDIQVLLDEERQAM
AKSRRLERSTNSFSYSSYHTLEEIYSWIDNFMVMEHSDIVSKIQIGNSFENQSILVLKFST
GGSRHPAIWIDTGIHSREWTHATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPD
GFAFTHSMNRLWRKNKSIIRPGIFCIGVDLNRNWKSGFGNGSNSNPCSETYHGFPSPQSEPE
VAAIVNFITAHGNFKALISIHYSQMLMYPYGRLLLEPVSNQRELYDLAKDAVEALYKVHG
IEYIFGSISTTLDVASGITVDWAYDSGIKYAFSFE LRDTGQYGFLLPATQIIPTAQETWM
ALRTIMEHTLNHPY

NOV97d

The disclosed NOV97d (alternatively referred to herein as CG57070-04) includes the 988 nucleotide sequence (SEQ ID NO:313) shown in Table 97G. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 973-975. The disclosed NOV97d maps to human chromosomes 7q31.

Table 97G. NOV97d Nucleotide Sequence (SEQ ID NO:313)

ATGCGGGGGTGTGCTGGTGTGAGTGTCCTGTTGGGGGGCTGCTTTGGCAAGGAGGACTTT
GTGGGGCATCAGGTGCTCCGAATCTCTGTAGCCGATGAGGCCAGGTACAGAAGGTGAAG
GAGCTGGAGGACCTGGAGCACCTGCAGCTGGACTTCTGGCGGGGGCTGCCACCCCTGGC
TCCCCCATCGACGTCCGAGTGCCCTTCCCCAGCATCCAGGCGGTCAAGATCTTCTGGAG
TCCCACGGCATCAGCTATGAGACCATGATCGAGGACGTGCAGTCGCTGCTGGACGAGGAG
CAGGAGCAGATGTTCCGCTTCCGGTCCCGGGCGCGCTCCACCGACACTTTTAACACGCC
ACCTACCACACCCTGGAGGAGGTGTATAGCTGGATTGACAACCTTTGTAATGGAGCATTC
GATATTGTCTCAAAAATTAGATTGGCAACAGCTTTGAAAACCAAGTCCATTCTTGTCTGT
AAGTTCAGCACTGGAGGTCTCGGCACCCAGCCATCTGGATTGACACTGGAATTCACCTCC
CGGGAGTGGATCACCATGCCACCGGCATCTGGACTGCCAATAAGATTGTCACTGATTAT
GGCAAAGACCGTGTCTGACAGACATACTGAATGCCATGGACATCTTCATAGAGCTCGTC
ACAAACCCTGATGGGTTTGTCTTTACCCACAGCATGAACCGCTTATGGCGGAAGAACAAG
TCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGGAAGTGAAGTCGGGT
TTTGGAGATGTGGCCAGTGGGATCACCGTCGACTGGGCCTACGACAGTGGCATCAAGTAC
GCCTTCAGCTTTGAGCTCCGGGACACTGGGCAGTATGGCTTCCTGCTGCCGGCCACACAG
ATCATCCCCACGGCCAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACACCCTG
AATCACCCCTACTAGCAGCAGCACTGAG

A NOV97d polypeptide (SEQ ID NO:314) encoded by SEQ ID NO:313 is 324 amino acids in length and is presented using the one-letter amino acid code in Table 97H. The Psort profile for NOV97d predicts that this sequence has a signal peptide and is likely to be localized to lysosomes with a certainty of 0.4757, or outside the cell with a certainty of

0.3703. The Signal P predicts a likely cleavage site for a NOV97d peptide is between positions 16 and 17, *i.e.*, at the dash in the sequence VFG-KE.

Table 97H. NOV97d Polypeptide Sequence (SEQ ID NO:314)
MRGLLVLSVLLGAVFGKEDFVGHQVLRISVADEAQVQKVKELEDLEHLQLDFWRGPAHPG SPIDVRVPFPSIQAVKIFLESHGISYETMIEDVQSLLDDEEQMFAFRSRARSTDTFNYA TYHTLEEVYSWIDNFVMEHSDIVSKIQIGNSFENQSIHLVKFSTGGSRHFAIWDITGIHS REWITHATGIWTANKIVSDYKDRVLTDLNAMDIFIELVTNPDGFAFTHSMNRLWRKNK SIRPGIFCIGVDLNRNWKSGFGDVASGITVDWAYDSGIKYAFSFELRDTGQYGFLLPATQ LIPTAQETWMALRTIMEHTLNHPY

5 *NOV97e*

The disclosed NOV97e (alternatively referred to herein as CG57070-05) includes the 1348 nucleotide sequence (SEQ ID NO:315) shown in Table 97I. A NOV97e ORF begins with a Kozak consensus ATG initiation codon at nucleotides 25-27 and ends with a stop codon at nucleotides 1333-1335. The disclosed NOV97e maps to human chromosome 7.

10

Table 97I. NOV97e Nucleotide Sequence (SEQ ID NO:315)
TGAAGCTCACCAGGAGGAAGAAGCATGCAGGGTACTCCTGGAGGCGGGACGCGCCCTGGG CCATCCCCCGTGGACAGGCGGACACTCCTGGTCTTCAGCTTTATCCTGGCAGCAGCTTTG GGCCAAATGAATTTTCACAGGGGACCAGGTTCTTCGAGTCTGGCCAAAGATGAGAAGCAG CTTTCAGTTCTCGGGGATCTGGAGGGCCTGAAACCCAGAAAGGTGGACTTCTGGCGTGGC CCAGCCAGGCCCAGCCTCCCTGTGGATATGAGAGTTCCCTTCTCCGAAGTGAAGACATC AAAGCTTATCTGGAGTCTCATGGACTTGCTTACAGCATCATGATAAAGGACATCCAGGTG CTGCTGGATGAGGAAAGACAGGCCATGGCGAAATCCCGCCGGCTGGAGCGCAGCACCAAC AGCTTCAGTTACTCATCATACCACACCCTGGAGGAGATATATAGCTGGATTGACAACTTT GTAATGGAGCATTCCGATATTGTCTCAAAAATTTCAGATTGGCAACAGCTTTGAAAACAG TCCATTCTTGCTCCTGAAGTTCAGCACTGGAGGTTCTCGGCACCCAGCCATCTGGATCGAC ACTGGAATTCAGTCCCGGGAGTGGATCAGCCATGCCACCGGCATCTGGACTGCCAATAAG ATTGTCAGTGATTATGGCAAAGACCGTGTCTGACAGACATACTGAATGCCATGGACATC TTCATAGAGCTCGTCACAAACCCTGATGGGTTTGCTTTTACCCACAGCATGAACCGCTTA TGGCGGAAGAACAAGTCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGG AACTGGAAGTCGGGTTTTGGAGGAAATGGTTCTAACAGCAACCCCTGCTCAGAACTTAT CACGGGCCCTCCCTCAGTCGGAGCCGGAGGTGGCTGCCATAGTGAACCTTCATCACAGCC CATGGCAACTTCAAGGCTCTGATCTCCATCCACAGCTACTCTCAGATGCTTATGTACCTT TACGCGCGATTGCTGGAGCCCGTTTCAAATCAGAGGGAGTTGTACGATCTTGCCAAGGAT GCGGTGGAGGCCTTGTATAAGGTCCATGGGATCGAGTACATTTTGGCAGCATCAGCACC ACCTCTATGTGGCCAGTGGGATCACCGTCGACTGGGCTATGACAGTGGCATCAAGTAC GCCTTCAGCTTTGAGCTCCGGGACACTGGGCAGTATGGCTTCTGCTGCCGGCCACACAG ATCATCCCCACGGCCAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACAACCTG AATCACCCCTACTAGCAGCAGCACTGAG

15 A NOV97e polypeptide (SEQ ID NO:316) encoded by SEQ ID NO:315 is 436 amino acids in length and is presented using the one-letter amino acid code in Table 97J. The Psort profile for NOV97e predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.5851. In alternative embodiments, a NOV97e polypeptide is located to lysosomes with a certainty of 0.4421. The Signal P predicts a likely

cleavage site for a NOV97e peptide is between positions 33 and 34, *i.e.*, at the dash in the sequence ALG-QM.

Table 97J. NOV97e Polypeptide Sequence (SEQ ID NO:316)

MQGTPGGGTRPGSPVDRRLLVFSFILAAALGQMNFTGDQVLRVLAKDEKQLSLLGDLE
GLKPQKVDFWRGPARPSLPVDMRVPFSELKDIKAYLESHGLAYSIMIKDIQVLLDEERQA
MAKSRRLERSTNSFSYSSYHTLEEIYSWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFS
TGGSRHPAIWIDTGIHSREWTHATGIWTANKIVSDYGKDRVLTDLNAMDIFIELVTNP
DGFAPTHSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPSQSE
PEVAAI VNFITAHGNFKALISIHYSQMLMYPYGRLLPEVSNQRELYDLAKDAVEALYKV
HGIEYIFGSISTTLYVASGITVDWAYDSGIKYAFSELRDGTGQYGFLLPATQIIPTAQET
WMALRTIMEHNLNHPY

5

NOV97f

The disclosed NOV97f (alternatively referred to herein as CG57070-06) includes the 975 nucleotide sequence (SEQ ID NO:317) shown in Table 97K. A SEC2 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TAG codon at nucleotides 973-975. The disclosed NOV97f maps to human chromosome 7q31.

10

Table 97K. NOV97f Nucleotide Sequence (SEQ ID NO:317)

ATGCGGGGGTGTGCTGGTGTGAGTGTCTGTTGGGGGGCTGTCTTTGGCAAGGAGGACTTT
GTGGGGCATCAGGTGCTCCGAATCTCTGTAGCCGATGAGGCCAGGTACAGAAGGTGAAG
GAGCTGGAGGACCTGGAGCACCTGCAGCTGGACTTCTGGCGGGGGCTGCCACCCTGGC
TCCCCCATCGACGTCCGAGTGCCCTTCCCAGCATCCAGGCGGTCAAGATCTTCTGGAG
TCCACGGCATCAGCTATGAGACCATGATCGAGGACGTGCAGTCGTGCTGGACGAGGAG
CAGGAGCAGATGTTCCGCTTCCGGTCCCGGGCGCGCTCCACCGACACTTTTAACTACGCC
ACCTACCACACCCTGGAGGAGGTGTATAGCTGGATTGACAACCTTTGTAATGGAGCATTCC
GATATTGTCTCAAATATTCAGATTGGCAACAGCTTTGAAAACCAAGTCCATTCTGTCTCTG
AAGTTCAGCACTGGAGGTTCTCGGCACCCAGCCATCTGGATCGACACTGGAATTCACCTC
CGGGAGTGGATCACCGTGCCACCGCATCTGGACTGCCAATAAGATTGTCTAGTGATTAT
GGCAAAGACCGTGTCTTGACAGACATACTGAATGCCATGGACATCTTCATAGGGCTCGTC
ACAAACCCTGATGGGTTTGTCTTTTACCCACAGCATGAACCGCTTATGGCGGAAGACAAG
TCCATCAGACCTGGAATCTTCTGCATCGGCGTGGATCTCAACAGGAAGTGAAGTCGGGT
TTTGGAGATGTGGCCAGTGGGATCACCGTCGACTGGGCCTATGACAGTGGCATCAAGTAC
GCCTTCAGCTTTGAGCTCCGGGACACTGGGCAGTATGGCTTCTGTGCGCGGCACACAG
ATCATCCCCACGGCCAGGAGACGTGGATGGCGCTTCGGACCATCATGGAGCACATCCTG
AATCACCCCTACTAG

15

A NOV97f polypeptide (SEQ ID NO:318) encoded by SEQ ID NO:317 is 324 amino acids in length and is presented using the one-letter amino acid code in Table 97L. The Psort profile for NOV97f predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.3989. In alternative embodiments, a NOV97f polypeptide is located to lysosomes with a certainty of 0.5061. The Signal P predicts a likely cleavage site for a NOV97f peptide is between positions 16 and 17, *i.e.*, at the dash in the sequence VFG-KE.

Table 97L. NOV97f Polypeptide Sequence (SEQ ID NO:318)

MRGLLVLSVLLGAVFGKEDFVGHQVLRISVADEAQVQVKELEDLEHLQLDFWRGPAHPG
SPIDVRVPFPSIQAVKIFLESHGISYETMIEDVQSLLEDEEQMFAPRSRARSTDTFNVA
TYHTLEEVYSWIDNFMVMEHSDIVSNIQIGNSFENQSILVLKFTSTGGSRHPAIIWIDTGIHS
REWITRATGIWTANKIVSDYGKDRVLTDLNAMDIFIGLVTNPDGFAFTHSMNRLWRKNK
SIRPGIFCIGVDLNRNWKSGFGDVASGITVDWAYDSGIKYAFSFE LRDTGQYGFLPATQ
IIPTAQETWMA LR TIMEHILNHPY

A BLAST analysis of NOV97 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV97 had high
5 homology to other proteins as shown in Table 97M.

Table 97M. BLASTX results from PatP database for NOV97

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P(N)
patp:AAE01663 Novel human protease #2	1898	9.3e-196
patp:AAB47565 Protease PRPS-7	1898	9.3e-196
patp:AAE01664 Novel human protease #3	1013	1.8e-174
patp:AAR97618 Human carboxypeptidase A1	1682	7.2e-173
patp:AAY28915 Human regulatory protein HRGP-1	1682	7.2e-173

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1017 of 1264 bases (80%) identical to a gb:GENBANK-
10 ID:HSPCBXA1|acc:X67318.1 mRNA from *Homo sapiens* (H.sapiens mRNA for procarboxypeptidase A1). The full amino acid sequence of the protein of the invention was found to have 315 of 419 amino acid residues (75%) identical to, and 357 of 419 amino acid residues (85%) similar to, the 419 amino acid residue ptrn:SWISSNEW-ACC:P15085 protein from *Homo sapiens* (Human) (CARBOXYPEPTIDASE A1 PRECURSOR (EC 3.4.17.1)).
15 NOV97 also has homology to the other proteins shown in the BLASTP data in Table 97N.

Table 97N. NOV97 BLASTP results

Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 17939400 ref NP_525124.1 (NM_080385)	carboxypeptidase A5 [Homo sapiens]	436	354/418 (84)	382/418 (90)	0.0
gi 4502997 ref NP_001859.1 (NM_001868)	pancreatic carboxypeptidase A1 precursor; Carboxypeptidase A [Homo sapiens]	419	315/419 (75)	357/419 (85)	0.0

gi 13528975 gb AAH05279.1 AAH05279 (BC005279)	carboxypeptidase A1 (pancreatic) [<i>Homo sapiens</i>]	419	314/419 (74)	356/419 (84)	0.0
gi 4336196 gb A AD17690.1 (AF076222)	carboxypeptidase A1 precursor [<i>Sus scrofa</i>]	419	295/419 (68)	345/419 (81)	e-177
gi 115878 sp P00730 CBPA_BOVIN	CARBOXYPEPTIDASE A PRECURSOR	419	289/419 (68)	345/419 (81)	e-176

This BLASTP data is displayed graphically in the ClustalW in Table 970. A multiple sequence alignment is given, with the NOV97a-f proteins being shown on lines 1-6 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 97N.

Table 970. ClustalW Alignment of NOV97	
NOV97a	(SEQ ID NO:308)
NOV97b	(SEQ ID NO:310)
NOV97c	(SEQ ID NO:312)
NOV97d	(SEQ ID NO:314)
NOV97e	(SEQ ID NO:316)
NOV97f	(SEQ ID NO:318)
gi 17939400	(SEQ ID NO:784)
gi 4502997	(SEQ ID NO:785)
gi 13528975	(SEQ ID NO:786)
gi 4336196	(SEQ ID NO:787)
gi 115878	(SEQ ID NO:788)
<div> <div> 1020304050 </div> <div> NOV97aMR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE NOV97bMR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE NOV97cMQGTPCGGTRPGSPVDRRTLLVFSFTLAAALGQMNFTGQVLRVRLAKDE NOV97dMR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE NOV97eMQGTPCGGTRPGSPVDRRTLLVFSFTLAAALGQMNFTGQVLRVRLAKDE NOV97fMR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE gi 17939400 MQGTPCGGTRPGSPVDRRTLLVFSFTLAAALGQMNFTGQVLRVRLAKDE gi 4502997 MR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE gi 13528975 MR---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE gi 4336196 MW---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVDE gi 115878 MQ---G-----LLVLSVLLGAVFGKEDFVGHQVLRISVADE </div> </div>	
<div> <div> 60708090100 </div> <div> NOV97aAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG NOV97bAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG NOV97cKQSLSLGLEGLKPKQVDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG NOV97dAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG NOV97eKQSLSLGLEGLKPKQVDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG NOV97fAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG gi 17939400 KQSLSLGLEGLKPKQVDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG gi 4502997 AQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG gi 13528975 AQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG gi 4336196 AQVQKVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG gi 115878 AEVQTVKELEDLEHLQLDFWRGPAHPGSPIDVRVPFPSIQAVKIFLESHG </div> </div>	
<div> <div> 110120130140150 </div> <div> NOV97aISYETMIEDVQS---LLDEEQEQMFAFRSRARSTDTFNYATYHTLEEY NOV97bLAYSIMIKDQVQVLLDEEQEQMAKSRRLERSTNSFSYSSYHTLEEY NOV97cLAYSIMIKDQVQVLLDEEQEQMAKSRRLERSTNSFSYSSYHTLEEY NOV97dISYETMIEDVQS---LLDEEQEQMFAFRSRARSTDTFNYATYHTLEEY NOV97eLAYSIMIKDQVQVLLDEEQEQMAKSRRLERSTNSFSYSSYHTLEEY </div> </div>	

NOV97f	ISYETMIEDVQS----	LLDEEQEQMFARSRARSTDTFNYATYHTLEEVI
gi 17939400	LAYSIMIKDIOV----	LLDEEQEQMAKSRRLERSTNSFSYSSYHTLEEVI
gi 4502997	ISYETMIEDVQS----	LLDEEQEQMFARSRARSTDTFNYATYHTLEEVI
gi 13528975	ISYETMIEDVQS----	LLDEEQEQMFARSRARSTDTFNYATYHTLEEVI
gi 4336196	IRYTIMIEDVQL----	LLDEEQEQMFASQGRARITSTFNYATYHTLEEVI
gi 115878	IRYRIMIEDVQS----	LLDEEQEQMFASQSRARSTNTFNYATYHTLEEVI

	160	170	180	190	200
NOV97a	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
NOV97b	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
NOV97c	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
NOV97d	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
NOV97e	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
NOV97f	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
gi 17939400	SWIDNFVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHPAIWIDTGIH				
gi 4502997	DFDILLVAENPHLVSKIQIGNTYEGRPITYVLKFSTGGSKRPAIWIDTGIH				
gi 13528975	DFDILLVAENPHLVSKIQIGNTYEGRPITYVLKFSTGGSKRPAIWIDTGIH				
gi 4336196	DFMDILVAEHPQLVSKIOIGSSYEGRPITYVLKFSTGGSNNRPAIWIDTGIH				
gi 115878	DFMDILLVAEHPQLVSKIOIGRSYEGRPITYVLKFSTGGSNNRPAIWIDTGIH				

	210	220	230	240	250
NOV97a	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
NOV97b	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
NOV97c	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
NOV97d	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
NOV97e	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
NOV97f	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
gi 17939400	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
gi 4502997	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
gi 13528975	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
gi 4336196	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				
gi 115878	SREWITHTATGIWTANKIVSDYGKDRVLTDLINAMDIFIELVTNPDGFAFT				

	260	270	280	290	300
NOV97a	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
NOV97b	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
NOV97c	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
NOV97d	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
NOV97e	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
NOV97f	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
gi 17939400	HSMNRLWRKNKSIRPGIFCIGVDLNRNWKSGFGGNGSNSNPCSETYHGPS				
gi 4502997	HSTNRMWRKTRSHTAGSLCIGVDPNRNWDAGFGLSGASSNPCSETYHCKF				
gi 13528975	HSTNRMWRKTRSHTAGSLCIGVDPNRNWDAGFGLSGASSNPCSETYHCKF				
gi 4336196	HSENRNWRKTRSRSGSFCVGVDPNRNWDAGFGGAGASSNPCSETYHCKF				
gi 115878	HSONRLWRKTRSVTSSSLCVGVDPNRNWDAGFGGAGASSNPCSETYHCKY				

	310	320	330	340	350
NOV97a	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
NOV97b	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
NOV97c	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
NOV97d	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
NOV97e	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
NOV97f	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
gi 17939400	PQSEPEVAAIIVNFITAHGNFKALISIHYSYSQMLMYPYGRLLPEVSNOREL				
gi 4502997	ANSEVEVKSIVDFVKDHGNIKAFISIHYSYSQMLMYPYGYKTEPVPDODEL				
gi 13528975	ANSEVEVKSIVDFVKDHGNIKAFISIHYSYSQMLMYPYGYKTEPVPDODEL				
gi 4336196	PNSEVEVKSIVDFVNDHGNIKAFISIHYSYSQMLMYPYGYKTEPAPADKDEL				
gi 115878	ANSEVEVKSIVDFVKDHGNIKAFISIHYSYSQMLMYPYGYTTQSTPDKTEL				

	360	370	380	390	400
NOV97a	YDLAKDAVEALYKVHGIETIYFGSISTTLVVASGITVDWAYDSGIKYAFSF				
NOV97b	YDLAKDAVEALYKVHGIETIYFGSISTTLVVASGITVDWAYDSGIKYAFSF				
NOV97c	YDLAKDAVEALYKVHGIETIYFGSISTTLVVASGITVDWAYDSGIKYAFSF				
NOV97d	-----DVASGITVDWAYDSGIKYAFSF				

NOV97e	YDLAKDAVEALYKVHGIIEYIFGSIISTTLVVASGITVDWAYDSGIKYAFSF
NOV97f	-----DVASGITVDWAYDSGIKYAFSF
gi 17939400	YDLAKDAVEALYKVHGIIEYIFGSIISTTLVVASGITVDWAYDSGIKYAFSF
gi 4502997	DQLSKAAVTALASLYCTKENYGSIIKAIYQASGSTLDWTYSQGIKYSFTF
gi 13528975	DQLSKAAVTALASLYCTKENYGSIIKAIYQASGSTLDWTYSQGIKYSFTF
gi 4336196	DQISKSAAVLTSLYCTKFOYGSIIITTIYQASGGTLDWTYNOGIKYSFSTF
gi 115878	NQVAKSAVEALKSLYCTSYKYGSIIITTIYQASGGSIDWSYNOGIKYSFTF

	410	420	430	440
NOV97a	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
NOV97b	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
NOV97c	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
NOV97d	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
NOV97e	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
NOV97f	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
gi 17939400	ELRDTGQYGFLLPATQIIPTAQETWMALRTIMEHTLNHPY			
gi 4502997	ELRDTGRYGFLLPASQIIPTAKETWALLTIMEHTLNHPY			
gi 13528975	ELRDTGRYGFLLPASQIIPTAKETWALLTIMEHTLNHPY			
gi 4336196	ELRDTGRYGFLLPASQIIPTAQETWALLTIMEHTLNHPY			
gi 115878	ELRDTGRYGFLLPASQIIPTAQETWALLTIMEHTLNHPY			

Table 97P lists the domain description from DOMAIN analysis results against NOV97. This indicates that the NOV97 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 97P. Domain Analysis of NOV97				
gnl Smart smart00631, Zn_pept, Zn_pept domain SEQ ID NO:887				
CD-Length = 286 residues, 100.0% aligned				
Score = 301 bits (770), Expect = 6e-83				
NOV97:	122	YHTLEEVS	WIDNFM	EHSDIVSKIQIGNSFENQ
Sbjct:	1	YHSYEEIE	AWLKKLA	ARYPDVRLVLSIGKSVEGRPIWVLKISNGPGRDGKPAVWIDAGIH
NOV97:	180	SREWIT	HATGIW	TANKIVSDY
Sbjct:	61	AREWIG	PATALY	LINQLLENYGSDFRVTKLLDKTDWYIVPVLNPDGYEYTHTSDRLLWRKN
NOV97:	240	KSIRPG	IFCIGV	DLNRNWKSGFGGNGSNSNPCSETYHGPSPQSEPEVAAIVNFITAHGNF
Sbjct:	121	RSPNSG	SNCRGV	DLNRNFPFWGETGASSNPCSETYAGSPFSEPETKAVRDFLRNKRKI
NOV97:	300	KALISIH	SYSQLM	YPYGRLL
Sbjct:	181	KLYIDL	HSYSQL	LILYPYGYTKNDLPPNVEDLPEVAKALADALASVHGGTRYTYGISNGAL
NOV97:	358	YVASGIT	VDWAYD	-SGIKYAFSFELRDTGQYGFLLPATQIIPTAQE
Sbjct:	241	YPASGG	SDDWAY	GTGTVLPFSYTTLELRDKGRYGFLLPSPQIIPTGWE

Carboxypeptidase A (EC 3.4.2.1) is a pancreatic exopeptidase. Three different forms of human pancreatic procarboxypeptidase A have been isolated. The A1 and A2 (600688) forms are monomeric proteins with different biochemical properties. Honey et al. (1984, 1986) found that an 8.6-kb human DNA fragment (detected by means of a rat cDNA probe for CPA) cosegregated with chromosome 7. The assignment was narrowed by demonstration of absence

of the human DNA fragment in cells with a deletion of 7q22-qter. By studying mouse-hamster hybrid cells, Honey et al. (1986) assigned the CPA gene to mouse chromosome 6. Trypsin (276000) is also on human 7q22-qter and on mouse 6. Stewart et al. (1990) concluded from multipoint linkage analysis with established chromosome 7 markers that the most likely location of carboxypeptidase is 7q31-qter. It lies distal to cystic fibrosis at a distance of approximately 12 cM.

NOV97 is predicted to be expressed in at least the following tissues: pancreas. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV97 is provided in Example 2.

The NOV97 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV97 nucleic acids encoding the carboxypeptidase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a carboxypeptidase A1-like protein includes the nucleic acid whose sequence is provided in Table 97A, 97C, 97E, 97G, 97I, or 97K or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 97A, 97C, 97E, 97G, 97I, or 97K while still encoding a protein that maintains its carboxypeptidase A1-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequences of Table 97A, 97C, 97E, 97G, 97I, or 97K, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least

in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 20% of the bases may be so changed.

5 The novel protein of the invention includes the carboxypeptidase A1-like protein whose sequence is provided in Table 97B, 97D, 97F, 97H, 97J, or 97L. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 97B, 97D, 97F, 97H, 97J, or 97L while still encoding a protein that maintains its carboxypeptidase A1-like activities and physiological functions, or a
10 functional fragment thereof. In the mutant or variant protein, up to about 25% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using
15 prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV98

20 The disclosed NOV98 (alternatively referred to herein as CG56939-01) includes the 5583 nucleotide sequence (SEQ ID NO:319) shown in Table 98A. A NOV98 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 4630-4632. The disclosed NOV98 maps to human chromosome 1.

Table 98A. NOV98 Nucleotide Sequence (SEQ ID NO:319)

```
GCCGCGTGCGTCTGCCCCGAGCCGGCGGGACATGCCCGGAGCGCGCTGGAGCGGCGC
GAGGAGGAGGCGAACGTGGTGCTACCGGGACGGTGGAGGAGATCCTCAACGTGGACCCG
GTGCAGCACACGTACTCTGCAAGGTTCCGGTCTGGCGGTACTTGAAGGGCAAAGACCTG
GTGGCCCGGAGAGCCTGCTGGACGGCGGCAACAAGTGGTGATCAGCGGCTTGGAGAC
CCCTCATCTGTGACAACCAAGGTGCCACTGGGGACACCAGGATCTTCTTTGTGAACCT
GCACCCCATACCTGTGGCCAGCCACAAGAACGAGCTGATGCTCAACTCCAGCCTCATG
CGGATCACCTGCGGAACCTGGAGGAGGTGGAGTTCTGTGTGGAAGATAAACCCGGGACC
CACTTCAGTCCAGTGCCCTCCGACGCTCCTGATGCGTGCCGGGAATGCTGTGCGGCTTC
GGCGCGGTGTGCGAGCCCAACGCGGAGGGGCGGGCGGGCGTCTGCGTCTGCAAGAAG
AGCCCGTGCGCCAGCGTGGTGCGCCTGTGTGTGGGTGCGACGCCTCCACCTACAGCAAC
GAATGCGAGCTGCAGCGGGCGCAGTGCCAGCCAGCAGCGCCGATCCGCTGCTCAGCCG
GGGCGGTGCGGCTCGCGGACCCCTGCTCCTCAACGTGACCTGCAGCTTCGGCAGCACCTGT
GCGCGCTCGGCCGACGGGTGACGGCCTCGTGCTGTGCCCCGCGACCTGCCGTGGCGCC
CCCGAGGGGACCGTCTGCGGCAGCGACGGCGCGGACTACCCCGCGAGTGCCAGCTCCTG
CGCCGCGCCTGCGCCCGCCAGGAGAATGTCTTCAAGAAGTTCGACGGCCCTTGTGACCC
TGTCAGGGCGCCCTCCCTGACCCGAGCCGAGCTGCCGTGTGAACCCGCGCACGCGCGC
CCTGAGATGCGCCTACGGCCCGAGAGCTGCCCTGCCCGGAGGCGCCAGTGTGTGGGGAC
GACGAGTACCTACGAAAACGACTGTGTGATGGGCCGATCGGGGGCGCCCGGGGTCTC
CTCCTGCAGAAAGTGCGTCCGGCCAGTGCCAGGGTCGAGACCAGTGCCCGGAGCCCTGC
```


AGGCCCTGCGAGGGGCTATCGAGAGGAGCTCACTGTGGGATGGGGTTGACCTCTGCCGCC
TGCCTGGGTATCTGGGCTGGCCATGGCTGTGTTCTTCATGTGTTGATTTTATTTGACCC
CTGGAGTGGTGGGTCTCATCTTTCCCATCTCGCCTGAGAGCGGCTGAGGGCTGCCTCACT
GCAAATCCTCCCCACAGCGTCAGTGAAAGTCGTCTTGTCTCAGAATGACCAGGGGCCAG
CCAGTGTCTGACCAAGGTCAAGGGGCAGGTGCAGAGGTGGCAGGGATGGCTCCGAAGCCA
GAA

A NOV98 polypeptide (SEQ ID NO:320) encoded by SEQ ID NO:319 is 1543 amino acids in length and is presented using the one-letter amino acid code in Table 98B. The Psort profile for NOV98 predicts that this sequence has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV98 polypeptide is located to microbodies with a certainty of 0.3000.

Table 98B. NOV98 Polypeptide Sequence (SEQ ID NO:320)

AACVLPAGAGGTCPERALERREEEENVVLTGTVEEILNVDPVQHTYSCKVRVWRYLKGKDL
VARESLLDGKNKVVISGFGDPLICDNQVSTGDTRIFFVNPAPPYLWPAHKNELMLNSSL
RITLRNLEEFVFCVEDKPGTHFTVPVPTPPDACRGMLCGFAGVCEPNAEGPGRASCVCCK
SPCPSVAVPCGSDASTYSNECELQRAQCSQRRIRLLSRGPCGRDPCSNVTSFGSTC
ARSADGLTASCLCPATCRGAPEGTVCGSDGADYPGECQLLRACARQENVFKKFDGPCDP
CQCALPDPSRSCRVPNRRPEMRLRPESCPARQAPVCGDDGVTYENDCVMGRSGAARGL
LLQKVRSGQCQGRDQCPEPCRFNAVCLSRGRPRCSCDRVTCDGAYRPVCAQDGRITYDSD
CWRQQAECRQRAIPSKHQGPCDQAPSPCLGVQCAFGATCAVKNQAACECLQACSSLYD
PVCSDGVITYGSACELEATACTLGREIQVARKGPCDRCGQCRFGALCEAETGRCVCPSEC
VALAQPVCGSDGHTYPSCEMLHVHACTHQISLHVASAGPCETCGDAVCAFGAVCSAGQCV
CPRCEHPPPGPVCGSDGVITYGSACELREAACLQQTQIEEARAGPCEQAECGSGSGSGED
GDCEQELCRQGGIWEDESDGPCVDFSCQSVPGSPVCGSDGVITYSTECELKKARCESQ
RGLYVAAQGACRGPTFAPLPPVAPLHCAQTPYGCCQDNITAARGVGLAGCPSACQCNPHG
SYGGTCDPATGQCSCRPGVGGRLCDRCPEPGFWNFRGI VTDGRSGCTPCSDPQGA VRDDC
EQMTGLCSCKPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGARCVEESGSA
HVCVPMLTCPANATKVCSDGVITYGNECQLKTIACRRCHLRQGLQISIQSLGCPQEA
VAVPVLTVPTAPSPAPSLVASAFGESGSTDGSSDEELSGDQASGGSGGPELEGSSVA
TPGPPVERASCYNPCHGAAPCRVLP EGGAQCEPLGREGTFCQTASGQDGS GPFPLADFNG
FSHLELRGLHTFARDLGEKMALEVFLARGPSGLLLYNGQKTGDKGDFVSLALRDRRLF
RYDLGKGA AVIRSREPVT LGAWTRVSLERNRKGALRVGDGPRVLGESPVHTVLNLKEP
LYVGGAPDFSKLAAAAVSSGFDGAIQLVSLGGRQLLTPEHVL RQVDVTSFAGHPCTRAS
GHPCLNGASCVPREAAVYVCLCPGFGSGPHCEKGLVEKSAGDVTDLAFDGRFTFVEY LNAVT
ESEKALQSNHFE LSLRTEATQGLVLWSGKATERADYVALAIVDGHQLSYNLG SQPVVLR
STVPVNTNRWLRVVAHREQREGSLQVGN EAPVTGSSPLGATQLD TDGALWLGGLPELPVG
PALPKAYGTGFVGCLRDVVVGRHPLHLEDAVTKPELRPCPTP

A BLAST analysis of NOV98 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV98 had high homology to other proteins as shown in Table 98C.

Table 98C. BLASTX results from PatP database for NOV98

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P(N)
patp:AAW26609 Human agrin - <i>Homo sapiens</i> , 492 aa.	2349	2.2e-246
patp:AAB93754 Human protein sequence	2179	1.5e-225

patp:AAV73993 Human prostate tumor EST fragment	2177	2.5e-225
patp:AAB31889 Amino acid sequence of a human protein	380	1.2e-62
patp:AAU16938 Human novel secreted protein	551	1.2e-51

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 3158 of 3416 bases (92%) identical to a gb:GENBANK-ID:AF016903|acc:AF016903.1 mRNA from *Homo sapiens* (agrin precursor mRNA). The full amino acid sequence of the protein of the invention was found to have 1092 of 1114 amino acid residues (98%) identical to, and 1093 of 1114 amino acid residues (98%) similar to, the 2026 amino acid residue ptnr:SPTREMBL-ACC:O00468 protein from *Homo sapiens* (Human) (AGRIN PRECURSOR). NOV98 also has homology to the other proteins shown in the BLASTP data in Table 98D.

10

Table 98D. NOV98 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 2988422 gb AAC39776.1 (AF016903)	agrin precursor [<i>Homo sapiens</i>]	2026	1088/1097 (99)	1088/1097 (99)	0.0
gi 399021 sp P25304 AGRI_RAT	Agrin precursor	1959	772/958 (80)	825/958 (85)	0.0
gi 202800 gb AAA40703.1 (M64780)	agrin [<i>Rattus norvegicus</i>]	1940	772/958 (80)	825/958 (83)	0.0
gi 202799 gb AAA40702.1 (M64780)	agrin [<i>Rattus norvegicus</i>]	1937	769/958 (80)	822/958 (85)	0.0
gi 399020 sp P31696 AGRI_CHICK	Agrin precursor	1169	637/1017 (62)	758/1017 (73)	0.0

This BLASTP data is displayed graphically in the ClustalW in Table 98E. A multiple sequence alignment is given, with the NOV98 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 98D.

15

Table 98E. ClustalW Alignment of NOV98	
NOV98	(SEQ ID NO:320)
gi 2988422	(SEQ ID NO:789)
gi 399021	(SEQ ID NO:790)
gi 202800	(SEQ ID NO:791)
gi 202799	(SEQ ID NO:792)
gi 399020	(SEQ ID NO:793)
<div> <div>1020304050</div> <div>..... </div> </div>	

NOV98 AACVLPGAGGTCPERALERREEEANVVLTGTVEEILNVDPVQHTYSCKVR
 gi|2988422| AACVLPGAGGTCPERALERREEEANVVLTGTVEEILNVDPVQHTYSCKVR
 gi|399021| -----MPPLPLEHRPROE-----
 gi|202800| -----MPPLPLEHRPROE-----
 gi|202799| -----MPPLPLEHRPROE-----
 gi|399020| -----

 60 70 80 90 100
 NOV98 VWRYLKCKDLVARESLLDGGNKVVISGFGDPLICDNQVSTGDRIFVNP
 gi|2988422| VWRYLKCKDLVARESLLDGGNKVVISGFGDPLICDNQVSTGDRIFVNP
 gi|399021| -----PGASMLVRY-----
 gi|202800| -----PGASMLVRY-----
 gi|202799| -----PGASMLVRY-----
 gi|399020| -----

 110 120 130 140 150
 NOV98 APPYLWPAHKNELMLNSSLRITLRNLEEVEFCVED-----KPGTHFT
 gi|2988422| APPYLWPAHKNELMLNSSLRITLRNLEEVEFCVED-----KPGTHFT
 gi|399021| ---FMIPCNICLILLATSTLGFAV-----LLFLSNY-----KPGTHFT
 gi|202800| ---FMIPCNICLILLATSTLGFAV-----LLFLSNY-----KPGTHFT
 gi|202799| ---FMIPCNICLILLATSTLGFAV-----LLFLSNY-----KPGTHFT
 gi|399020| ---MPPAHRNELMLNSSLRITLRNLEEVEHCVEEHRLKLLADKPNSTFT

 160 170 180 190 200
 NOV98 PVPPPTPPDACRGMLCGFGAVCEPNAEGPGRASCVCCKSPCPSSVVPVCGS
 gi|2988422| PVPPPTPPDACRGMLCGFGAVCEPNAEGPGRASCVCCKSPCPSSVVPVCGS
 gi|399021| PAPPTPPDVCRGMLCGFGAVCEPSEVEDPGRASCVCCKKNACPATVAPVCGS
 gi|202800| PAPPTPPDVCRGMLCGFGAVCEPSEVEDPGRASCVCCKKNACPATVAPVCGS
 gi|202799| PAPPTPPDVCRGMLCGFGAVCEPSEVEDPGRASCVCCKKNACPATVAPVCGS
 gi|399020| QTPPTPPDACRGMLCGFGAVCERSPTDESPASCVCCKKTACPVVVPVCGS

 210 220 230 240 250
 NOV98 DASTYSNECELQRAQCSQORRIRLLSRGPCGSRDPCSNVTCFSGSTCARS
 gi|2988422| DASTYSNECELQRAQCSQORRIRLLSRGPCGSRDPCSNVTCFSGSTCARS
 gi|399021| DASTYSNECELQRAQCNOQRRIRLLRQGPCGSRDPCANVTCFSGSTCVPS
 gi|202800| DASTYSNECELQRAQCNOQRRIRLLRQGPCGSRDPCANVTCFSGSTCVPS
 gi|202799| DASTYSNECELQRAQCNOQRRIRLLRQGPCGSRDPCANVTCFSGSTCVPS
 gi|399020| DYSTYSNECELEKRAQCNOQRRIRKVISKGPCGSKDPCAEVTCFSGSTCVRS

 260 270 280 290 300
 NOV98 ADGLTASCLCPATCRGAPEGTVCSDGADYPGECQLLRACARQENVFKK
 gi|2988422| ADGLTASCLCPATCRGAPEGTVCSDGADYPGECQLLRACARQENVFKK
 gi|399021| ADGQTASCLCPPTTCFGAPDGTVCSDGVDYPSECQLLSHACASQEHIFKK
 gi|202800| ADGQTASCLCPPTTCFGAPDGTVCSDGVDYPSECQLLSHACASQEHIFKK
 gi|202799| ADGQTASCLCPPTTCFGAPDGTVCSDGVDYPSECQLLSHACASQEHIFKK
 gi|399020| ADGQTASCLCPASCSGVAESTVCSDGKDYRSECDLNKHACDKQENVFKK

 310 320 330 340 350
 NOV98 FDGPCDPCQGALPDPSRSRVRNPRTTRRPEMLLRPESCPARQAPVCGDDGV
 gi|2988422| FDGPCDPCQGALPDPSRSRVRNPRTTRRPEMLLRPESCPARQAPVCGDDGV
 gi|399021| FNGPCDPCQGSMSDLNHLICRVNPRTRHPEMLLRPENCPAHTPICGDDGV
 gi|202800| FNGPCDPCQGSMSDLNHLICRVNPRTRHPEMLLRPENCPAHTPICGDDGV
 gi|202799| FNGPCDPCQGSMSDLNHLICRVNPRTRHPEMLLRPENCPAHTPICGDDGV
 gi|399020| FDGACDPCRGILNDVNRVCRVNPRTRRVELLSPENCPSKREPVCDDGV

 360 370 380 390 400
 NOV98 TYENDCVMSRSGAARGLLQKVRSGQCQGRDQCEPCRFNAVCLSRGRGP
 gi|2988422| TYENDCVMSRSGAARGLLQKVRSGQCQGRDQCEPCRFNAVCLSRGRGP
 gi|399021| TYENDCVMSRIGATRGLLLQKVRSGQCQTRDQCEPCCFNSVCLSRGRGP
 gi|202800| TYENDCVMSRIGATRGLLLQKVRSGQCQTRDQCEPCCFNSVCLSRGRGP
 gi|202799| TYENDCVMSRIGATRGLLLQKVRSGQCQTRDQCEPCCFNSVCLSRGRGP
 gi|399020| TYASECVMSRTGAIRGLEHLQKVRSGQCQHODKCKDECFNAVCLRF-WHA

	410	420	430	440	450
NOV98	RCSCDRVTC	CGAYRPVCA	QDGRITYD	SDCWRRQA	ECRQQRAI
gi 2988422	RCSCDRVTC	CGAYRPVCA	QDGRITYD	SDCWRRQA	ECRQQRAI
gi 399021	HCSCDRVTC	CGSYRPVCA	QDGHITYN	NDCWRRQA	ECRQQRAI
gi 202800	HCSCDRVTC	CGSYRPVCA	QDGHITYN	NDCWRRQA	ECRQQRAI
gi 202799	HCSCDRV---	GSYRPVCA	QDGHITYN	NDCWRRQA	ECRQQRAI
gi 399020	RCSCDRITC	DGYRPVCA	RDSTRYSN	DCERQKAE	CHOKAATP

	460	470	480	490	500
NOV98	Q-APSPCLG	VQCAFGAT	CAVKNQA	ACECLQAC	SSLYDPV
gi 2988422	Q-APSPCLG	VQCAFGAT	CAVKNQA	ACECLQAC	SSLYDPV
gi 399021	Q-TPSPCHG	VQCAFGAV	CTVKNKA	ABCECOR	VCSGIYD
gi 202800	Q-TPSPCHG	VQCAFGAV	CTVKNKA	ABCECOR	VCSGIYD
gi 202799	Q-TPSPCHG	VQCAFGAV	CTVKNKA	ABCECOR	VCSGIYD
gi 399020	LGTPSPCL	SVETFGAT	CVVKNRE	PVCECQVC	QGRYDPV

	510	520	530	540	550
NOV98	ACELEATA	CTLGREIQ	VARRGPCD	RCGQCRFG	ALCEAETG
gi 2988422	ACELEATA	CTLGREIQ	VARRGPCD	RCGQCRFG	ALCEAETG
gi 399021	VCELESM	ACTLGREI	QVARRG	PCDPCGQ	CRFGLCE
gi 202800	VCELESM	ACTLGREI	QVARRG	PCDPCGQ	CRFGLCE
gi 202799	VCELESM	ACTLGREI	QVARRG	PCDPCGQ	CRFGLCE
gi 399020	PCELNMA	CVLKR	IRVKKH	RGPCDR	CGKCGF

	560	570	580	590	600
NOV98	LAQPVCGS	DGHTYF	SECM	LHVHACT	HQISLHV
gi 2988422	LAQPVCGS	DGHTYF	SECM	LHVHACT	HQISLHV
gi 399021	SAQPVCGS	DGHTYF	ASECEL	LHVHACT	HQISLHV
gi 202800	SAQPVCGS	DGHTYF	ASECEL	LHVHACT	HQISLHV
gi 202799	SAQPVCGS	DGHTYF	ASECEL	LHVHACT	HQISLHV
gi 399020	SSQPVCGT	DGNTYF	GSECEL	LHVRACT	QKNILVA

	610	620	630	640	650
NOV98	VCSAGQC	VCPRCEH	PPPGP	VCSDGVTY	CSACEL
gi 2988422	VCSAGQC	VCPRCEH	PPPGP	VCSDGVTY	CSACEL
gi 399021	VCSAGQC	VCPRCEH	PPPGP	VCSDGVTY	CSACEL
gi 202800	VCSAGQC	VCPRCEH	PPPGP	VCSDGVTY	CSACEL
gi 202799	VCSAGQC	VCPRCEH	PPPGP	VCSDGVTY	CSACEL
gi 399020	TCVGGQ	VCPRCE	QQQLA	QVCGT	DGLTYD

	660	670	680	690	700
NOV98	GPCEQAE	CGSGG	SGSGED	GDCEQEL	CRQRGGI
gi 2988422	GPCEQAE	CGSGG	SGSGED	GDCEQEL	CRQRGGI
gi 399021	GPCEPAE	CGSGG	SGSGED	DECEQEL	CRQRGGI
gi 202800	GPCEPAE	CGSGG	SGSGED	DECEQEL	CRQRGGI
gi 202799	GPCEPAE	CGSGG	SGSGED	DECEQEL	CRQRGGI
gi 399020	GPCED-	ECGSGG	SGSGED	GECEQEL	CRHYGG

	710	720	730	740	750
NOV98	VPGSPVCG	SDGVTY	STEC	LKKARCES	QQLYVAA
gi 2988422	VPGSPVCG	SDGVTY	STEC	LKKARCES	QQLYVAA
gi 399021	VPRSPVCG	SDGVTY	GTECDL	LKKARCES	QQLYVAA
gi 202800	VPRSPVCG	SDGVTY	GTECDL	LKKARCES	QQLYVAA
gi 202799	VPRSPVCG	SDGVTY	GTECDL	LKKARCES	QQLYVAA
gi 399020	VPRSPVCG	SDGVTY	ANEC	LKKTRCE	KRONLYV

	760	770	780	790	800
NOV98	APLHCAQ	TPYGCCQ	DNITAA	RGVGLAG	CPSACQ
gi 2988422	APLHCAQ	TPYGCCQ	DNITAA	RGVGLAG	CPSACQ
gi 399021	APFHCAQ	TPYGCCQ	DNITAA	RGVGLAG	CPSACH

gi	202800	AFPHCAQTPYGCCQDNFTAAQGVGLAGCPSTCHCNPHGSYS	GTCDPATGO
gi	202799	AFPHCAQTPYGCCQDNFTAAQGVGLAGCPSTCHCNPHGSYS	GTCDPATGO
gi	399020	PVVHCSQTIYGCCPDNMTLALGVGAAGCPSTCCCNPHGSYS	GGTCDPATGO
		810 820 830 840 850	
NOV98		CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
gi	2988422	CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
gi	399021	CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
gi	202800	CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
gi	202799	CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
gi	399020	CSCRPGVGGLRCDRCEPGFWNFRGIVTDGRSGCTPCSCDPRGAVRDDCEQ	
		860 870 880 890 900	
NOV98		MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
gi	2988422	MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
gi	399021	MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
gi	202800	MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
gi	202799	MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
gi	399020	MTGLCSCRPGVAGPKCGQCPDGRALGPAGCEADASAPATCAEMRCEFGAR	
		910 920 930 940 950	
NOV98		CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
gi	2988422	CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
gi	399021	CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
gi	202800	CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
gi	202799	CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
gi	399020	CVEESGSAHCVCPMLTCEANATKVCSDGVITYGNECQLKTIACRRCHLR	
		960 970 980 990 1000	
NOV98		QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
gi	2988422	QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
gi	399021	QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
gi	202800	QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
gi	202799	QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
gi	399020	QGLQISTQSLGPCQEAAPSTHPTSASVT-----VTTPEGLLLSQALP	
		1010 1020 1030 1040 1050	
NOV98		AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
gi	2988422	AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
gi	399021	AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
gi	202800	AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
gi	202799	AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
gi	399020	AP----PGALPLAPSSSTAHSQTTPPPSSRRPRTTASVPRITTVWPVLTVPPT	
		1060 1070 1080 1090 1100	
NOV98		APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
gi	2988422	APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
gi	399021	APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
gi	202800	APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
gi	202799	APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
gi	399020	APS-PAPSLVASAFGESGSTDGSSDE--ELSGDQEAASGGGSGGLEPPLGGS	
		1110 1120 1130 1140 1150	
NOV98		SVATPGPPVERASCYN-----	
gi	2988422	SVATPGPPVERASCYN-----	
gi	399021	SVATPGPPVERASCYN-----	
gi	202800	SVATPGPPVERASCYN-----	
gi	202799	SVATPGPPVERASCYN-----	
gi	399020	SVATPGPPVERASCYN-----	
		1160 1170 1180 1190 1200	

NOV98
gi|2988422| GVEGQELFYTPEMADPKSELFGETARSIESLDDLFNRNSDVKKDFRSVRL
gi|399021| GVEGQELFYTPEMADPKSELFGETARSIESLDDLFNRNSDVKKDFWSVRL
gi|202800| GVEGQELFYTPEMADPKSELFGETARSIESLDDLFNRNSDVKKDFWSVRL
gi|202799| GVEGQELFYTPEMADPKSELFGETARSIESLDDLFNRNSDVKKDFWSVRL
gi|399020| EVEGQELFYTPEMADPKSELFGETARSIESALDELFRNSDVKNDFKSRV

1210 1220 1230 1240 1250
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| RDLGPGKSVRAIVDVHFDPTTAFAFAPDVARALLRQIQVSRRRSLGVRRLPL
gi|399021| RELGPGKLVRAIVDVHFDPTTAFAFASDVGOALLRQIQVSRPWALAVRRPL
gi|202800| RELGPGKLVRAIVDVHFDPTTAFAFASDVGOALLRQIQVSRPWALAVRRPL
gi|202799| RELGPGKLVRAIVDVHFDPTTAFAFASDVGOALLRQIQVSRPWALAVRRPL
gi|399020| RDLGOSSAVRIVESHFDEATSYTAADVQAASLKQIRASCKRTILVKKPQ

1260 1270 1280 1290 1300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| QEHVRFMDFDWFPAFITGATSCAIAACATARATTASRLPSSAVTPRAPHP
gi|399021| QEHVRFMDFDWFPTFFTGAAATGTTAAMATARATTVSRLPASSVTTPR-VYP
gi|202800| QEHVRFMDFDWFPTFFTGAAATGTTAAMATARATTVSRLPASSVTTPR-VYP
gi|202799| QEHVRFMDFDWFPTFFTGAAATGTTAAMATARATTVSRLPASSVTTPR-VYP
gi|399020| QEHVRFMDFDWTERIFTTITTTTATTMAPATRRHTTASAAATTAHLIRQ

1310 1320 1330 1340 1350
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| SHTSOPVAKTTAAPTTRRP---PTTAPSRVPGRFPFAPQOEKPCDSQP
gi|399021| SHTSRPVGRITTAAPTTRRP---PTTATN-MDRPRTPGHQQPSKSCDSQP
gi|202800| SHTSRPVGRITTAAPTTRRP---PTTATN-MDRPRTPGHQQPSKSCDSQP
gi|202799| SHTSRPVGRITTAAPTTRRP---PTTATN-MDRPRTPGHQQPSKSCDSQP
gi|399020| DTVGHPKSAKLAPASTTRRPSTLPTTARRKPTRQPPSTTKKPSRPCDSQP

1360 1370 1380 1390 1400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| CFHGGTCQDQWALCGGFTCSCTAGRGGAUCEKVLGAPVPAFEGRSFLAFPT
gi|399021| CLHGGTCQDQDSCKGFTCSCTAGRGGSUCEKVQPPSMPAFKCHSFLAFPT
gi|202800| CLHGGTCQDQDSCKGFTCSCTAGRGGSUCEKVQPPSMPAFKCHSFLAFPT
gi|202799| CLHGGTCQDQDSCKGFTCSCTAGRGGSUCEKVQPPSMPAFKCHSFLAFPT
gi|399020| CLHGGTCQEDD--GREFTCRCPAGKGGAUCEKPIRYFTIPSEFGKSYLAFKM

1410 1420 1430 1440 1450
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| LRAYHTLRLEALEFRALEPOGLLLYNGNARGKDFLALALLDGRVQFRFDTG
gi|399021| LRAYHTLRLEALEFRALETEGLLLYNGNARGKDFLALALLDGRVQFRFDTG
gi|202800| LRAYHTLRLEALEFRALETEGLLLYNGNARGKDFLALALLDGRVQFRFDTG
gi|202799| LRAYHTLRLEALEFRALETEGLLLYNGNARGKDFLALALLDGRVQFRFDTG
gi|399020| MKAYHTVRIAMEFRATELSGLLLYNGNARGKDFISLALVGGFVELRFNTG

1460 1470 1480 1490 1500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| SGPAVLTSAPVPEPGQWHRLELSRHWRRGTLSDVDGETPVVGESPSGTDGL
gi|399021| SGPAVLTSAPVPEPGQWHRLELSRHWRRGTLSDVDGETPVVGESPSGTDGL
gi|202800| SGPAVLTSAPVPEPGQWHRLELSRHWRRGTLSDVDGETPVVGESPSGTDGL
gi|202799| SGPAVLTSAPVPEPGQWHRLELSRHWRRGTLSDVDGETPVVGESPSGTDGL
gi|399020| SCTGVITSKVRVEPGKWHOLVNNRNRSCMLAVDG-EHVSSESPTGTDGL

1510 1520 1530 1540 1550
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV98
gi|2988422| NLDTDLEFVGVPEDQAAVALERTFVGAGLRGCIRLLDYNQORLELGIGPG
gi|399021| NLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQOLELSDWQR
gi|202800| NLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQOLELSDWQR
gi|202799| NLDTNLYVGGIPEEQVAMVLDRTSVGVGLKGCIRMLDINNQOLELSDWQR
gi|399020| NLDTDLEFVGAPEDQAAVVAERTATVGLKGSIRLLDYNQMYLDRKGS

		1560	1570	1580	1590	1600
NOV98					
gi 2988422	AA	TRGSGVGECGDHPCLNPCHGGAP	CONLEAGR	FHCQC	PPGRVG	PTCAD
gi 399021	AA	VQSSGVGECGDHPCLNPCHGGAL	COALEAGM	FLCQC	PPGRFG	PTCAD
gi 202800	AA	VQSSGVGECGDHPCLNPCHGGAL	COALEAGM	FLCQC	PPGRFG	PTCAD
gi 202799	AA	VQSSGVGECGDHPCLNPCHGGAL	COALEAGM	FLCQC	PPGRFG	PTCAD
gi 399020	DV	LYGSGVGECGNDPCHPNCHHGAS	SCHVKEA	EMFHC	ECLHSY	TGPTCAD

		1610	1620	1630	1640	1650
NOV98		----- ----- ----- ----- ----- -----				
gi 2988422	EK	SPCQPNPCHGAAPCRVL	PEGGAQCE	CPLGRE	GTFCTAS	QDGS
gi 399021	EK	SPCQPNPCHGAAPCRVL	SSGGARCE	CPLGRSG	TFCQTV	LETAGSR
gi 202800	EK	SPCQPNPCHGAAPCRVL	SSGGARCE	CPLGRSG	TFCQTV	LETAGSR
gi 202799	EK	SPCQPNPCHGAAPCRVL	SSGGARCE	CPLGRSG	TFCQTV	LETAGSR
gi 399020	ER	NPCDPTPCHISATCLVL	PEGGAMC	ACPMGE	GEFCER	VTEQDHTMP

		1660	1670	1680	1690	1700
NOV98					
gi 2988422	AD	FNGFSHLELRGLHTF	ARDLGEK	MALEV	VFLARG	PSGLLLYNGQ
gi 399021	AD	FNGFSHLELRGLHTF	ARDLGEK	MALEV	VFLARG	PSGLLLYNGQ
gi 202800	AD	FNGFSHLELRGLHTF	ARDLGEK	MALEV	VFLARG	PSGLLLYNGQ
gi 202799	AD	FNGFSHLELRGLHTF	ARDLGEK	MALEV	VFLARG	PSGLLLYNGQ
gi 399020	PE	FNGFSYLELNLGLT	LF LTC	-RQMS	MEVVFL	AKSPSGMIF

		1710	1720	1730	1740	1750
NOV98					
gi 2988422	GD	FVSLALRDRLEFR	YDLGKGA	AVIRSRE	PVTLGA	WTRVSLERN
gi 399021	GD	FVSLALRDRLEFR	YDLGKGA	AVIRSRE	PVTLGA	WTRVSLERN
gi 202800	GD	FVSLALHNRHLEF	CYDLGKGA	AVIRSK	EPALGT	WVRVFLERN
gi 202799	GD	FVSLALHNRHLEF	CYDLGKGA	AVIRSK	EPALGT	WVRVFLERN
gi 399020	GD	FVSLALHNRHLEF	CYDLGKGA	AVIRSK	EPALGT	WVRVFLERN

		1760	1770	1780	1790	1800
NOV98					
gi 2988422	LR	VGDGPRVLGESP	----VP	HTVLNL	KEPLYV	GGAPDFSKL
gi 399021	LR	VGDGPRVLGESP	----VP	HTVLNL	KEPLYV	GGAPDFSKL
gi 202800	LO	VGDGPRVLGES	PKSRKV	PHTMLN	LKEPLY	IGGAPDFSK
gi 202799	LO	VGDGPRVLGES	PKSRKV	PHTMLN	LKEPLY	IGGAPDFSK
gi 399020	MR	INNGBRVMGES	PKSRKV	PHAFNL	NLKEE	FVVGAPDFSK

		1810	1820	1830	1840	1850
NOV98					
gi 2988422	FD	GAIQLVSLGGR	QLLTPE	HVLRQ	VDVTS	FAHPCTR
gi 399021	FD	GAIQLVSLGGR	QLLTPE	HVLRQ	VDVTS	FAHPCTR
gi 202800	FS	GVIQLVSLRGH	QLLTQ	EHVLR	AVDVSP	FADHPCT
gi 202799	FS	GVIQLVSLRGH	QLLTQ	EHVLR	AVDVSP	FADHPCT
gi 399020	FY	GAVQRISIK	GVPLLK	EHRS	AVEIST	FAHPCTQ

		1860	1870	1880	1890	1900
NOV98					
gi 2988422	PR	EAAVCLCPGG	FSGLH	CEKGL	VEKSA	GDVDTL
gi 399021	PR	EAAVCLCPGG	FSGLH	CEKGL	VEKSA	GDVDTL
gi 202800	PR	EATYECLCP	GGFSGLH	CEKGL	VEKSV	GDLET
gi 202799	PR	EATYECLCP	GGFSGLH	CEKGL	VEKSV	GDLET
gi 399020	PR	LESYECAC	QRFSCA	HCEK	VITEKA	AGDAE

		1910	1920	1930	1940	1950
NOV98		----- ----- ----- ----- -----				
gi 2988422	SE	KALQSNH	FELSLR	TEATQ	GLVLW	SGKATE
gi 399021	SE	LNEIPAP	ETLDS	RALFSE	KALQSNH	FELSLR

synapses during development might also help to understand mechanisms that cause changes in synapses during learning and memory.

For practical reasons, most of the current knowledge of synapse development is derived from studies of the vertebrate neuromuscular junction. Upon arrival of a motor axon at the muscle fiber, signals released from its growth cone initiate the formation of a synapse. This process consists of two stages: arrest of axon growth at the target area and differentiation of pre- and postsynaptic cells at the site of nerve-muscle contact.

Studies of regenerating neuromuscular junctions in vertebrates have revealed that important signals for the formation of this synapse are located in the synaptic basal lamina, and attempts to identify these signals have led to the isolation of agrin and other components. The induction of the intensively studied synapse between nerve and muscle is initiated by the binding of neuron-specific isoforms of the basal membrane protein agrin to receptors on the surface of myotubes. Agrin activates a receptor complex that includes the muscle-specific kinase and most likely additional, yet to be identified, components. Receptor activation leads to the aggregation of acetylcholine receptors (AChR) and other proteins of the postsynaptic apparatus. This activation process has unique features which distinguish it from other receptor tyrosine kinases. In particular, the autophosphorylation of the kinase domain, which usually induces the recruitment of adaptor and signalling molecules, is not sufficient for AChR aggregation. Apparently, interactions of the extracellular domain with unknown components are also required for this process.

Agrin binds to a second protein complex on the muscle surface known as the dystrophin-associated glycoprotein complex. This binding forms one end of a molecular link between the extracellular matrix and the cytoskeleton.

While many components of the machinery triggering postsynaptic differentiation have now been identified, the picture of the molecular pathway causing the redistribution of synaptic proteins is still incomplete. Recent advances implicate proteins such as dystroglycan, MuSK, and rapsyn in the transduction of agrin signals. Additional functions of agrin have been discovered, including the upregulation of gene transcription in myonuclei and the control of presynaptic differentiation.

Agrin therefore appears to play a unique role in controlling synaptic differentiation on both sides of the neuromuscular junction.

NOV98 is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart,

kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus, bone, cerebral medulla/ cerebral white matter, cervix, colon, epidermis, foreskin, hair follicles, liver, lung, lymphoid tissue, ovary, parathyroid gland, parietal lobe, retina, skin, vein, whole organism. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV98 is provided in Example 2.

The NOV98 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV98 nucleic acids encoding the agrin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a agrin-like protein includes the nucleic acid whose sequence is provided in Table 98B, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 98B while still encoding a protein that maintains its agrin-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 98B, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In the mutant or variant nucleic acids, and their complements, up to about 8% of the bases may be so changed.

The novel protein of the invention includes the agrin-like protein whose sequence is provided in Table 98B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 98B while still encoding a protein that maintains its agrin-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 2% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOV99

The disclosed NOV99 (alternatively referred to herein as CG57010-01) includes the 1513 nucleotide sequence (SEQ ID NO:321) shown in Table 99A. A NOV99 ORF begins with a Kozak consensus ATG initiation codon at nucleotides 396-398 and ends with a TGA codon at nucleotides 1410-1412. The disclosed NOV99 maps to human chromosome 14q32.33.

Table 99A. NOV99 Nucleotide Sequence (SEQ ID NO:321)

```
GGAGCCCCCGCCCTGGGATTCCCAGGTGTTTTTCATTTGGTGATCAGCACTGAACACAGAA
GAGTCATGACGGAGTTTGGGCTGAGCTGGGTTTTCCTTGTTGCTATTTTTAAAGGTGTCC
AGTGTGAGGTGCAGCTGGTGGAGTCTGGGGGAGACTTGGTCCAGCCTGGGGGCTCCCTGA
GACTCTCCTGTGCAGCCTCTGGATTACCTTCAGTAGTTATGCTATGCACCTGGGTCCGCC
AGGCTCCAGGGAAGGGACTGAAATATGTTTCAGGTATTAGTAGTAATGGGCGTAGAACAT
ATTATGCAAATCTGTGAAGGGCAGATTACCATCTCCAGAGACAATCCAAGAACACGT
TGTATCTTCAAATGGGCAGCCTGAGAGCTGAGGACATGGCTGTGTATTACTGTGTGTCGG
GGGGAATCTATGATAGTAGTGGTCCCTTTGACTACTGGGGCCAGGGAACCTGTCACCG
TCTCCTCAGCATCCCCGACCAGCCCCAAGGTCTTCCCGCTGAGCCTCTGCAGACCCAGC
CAGATGGGAACGTGGTTCATCGCCTGCCTGGTCCAGGGCTTCTTCCCCAGGAGCCACTCA
GTGTGACCTGGAGCGAAAGCGGACAGGGCGTGACCGCCAGAACTTCCACCCAGCCAGG
ATGCTCCGGGACCTGTACACCAGCAGCAGCCAGCTGACCTGCGCGCCACACAGTGCC
TAGCCGGCAAGTCCGTGACATGCCACGTGAAGCACTACACGAATCCAGCCAGGATGTGA
CTGTGCCCTGCCAGTTCCTCAACTCCACCTACCCCATCTCCCTCAACTCCACCTACCC
CATCTCCCTCATGTGCCACCCCGACTGTCTACTGCACCGACCGGCCCTCGAGGACCTGC
TCTTAGGTTTCAAGCGAACCCTACGTGCACACTGACCGGCTGAGAGATGCCTCAGGTG
TCACCTTCACCTGGACGCCCTCAAGTGGGAAGAGCGCTGTCAAGGACCACCTGAGCGTG
ACCTCTGTGGCTGCTACAGCGTGTCCAGTGTCTGCCTGGCTGTGCCAGCCATGGAACC
ATGGGGAGACCTTCACCTGCACTGTGCCACCCGAGTTGAAGACCCCACTAGTTCGCT
GGCTGCAGGGGTACAGGAGCTGCCCGCGAGAAGTACCTGACTTGGGCATCCCGGCAGG
AGCCAGCCAGGGCACCACCTTCGCTGTGACCAGCATACTGCGCGTGGCAGCCGAGG
ACTGGAAGAAGGGGACACCTTCTCTGCATGGTGGGCCACGAGGCCCTGCCGTGGCCT
TCACACAGAAGACCATCGACCGCTTGGCGGGTAAACCCACCCATGTCAATGTGTCTGTTG
```

TCATGGTGGAGGTGGACGGCACCTGCTACTGAGCCGCCCGCTGTCCCCACCCCTGAATA
 AACTCCATGCTCCCCCAA
 AAAAAAAAAA

A NOV99 polypeptide (SEQ ID NO:322) encoded by SEQ ID NO:321 is 338 amino acids in length and is presented using the one-letter amino acid code in Table 99B. The Psort profile for NOV99 predicts that this sequence has no signal peptide and is likely to be localized to the cytoplasm with a certainty of 0.4500. In alternative embodiments, a NOV99 polypeptide is located to microbodies with a certainty of 0.1315.

Table 99B. NOV99 Polypeptide Sequence (SEQ ID NO:322)

MAVYYCVSGGIYDSSGPFDDYWGQGTLLTVSSASPTSPKVFPLSLCSTQPDGNVVIACLVQ
 GFFPQEPLSVTWSESGQGVARNFPSPQDASGDLYTSSQLTLPATQCLAGKSVTCHVKH
 YTNPSQDVTVPFCVPSTPPTSPSTPPTSPSCCHPRLSLHRPALEDLLLGSEANLTCTL
 TGLRDASGVTFTWTTPSSGKSAVQGPPERDLGCGYSVSSVLPGCAQFPWNHGETFTCTAAHP
 ELKTPPLVRWLQGSQELPREKYLTVASRQEPSQGTTFFAVTSILRVAEDWKKGDTFSCMV
 GHEALPLAFTQKTIDRLAGKPTHVNVSVVMVEVDGTCY

A BLAST analysis of NOV99 was run against the proprietary PatP GENESEQ Protein Patent database. It was found, for example, that the amino acid sequence of NOV99 had high homology to other proteins as shown in Table 99C.

Table 99C. BLASTX results from PatP database for NOV99

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum
		Probability P (N)
patp:AA88483 Cancer suppressor gene product	1335	9.8e-187
patp:AAB82914 Human immune response protein HIRP3	1268	1.2e-179
patp:AAM93283 Human polypeptide,	1266	1.9e-179
patp:AA44723 Human immune system molecule, ISMO-4	1262	5.1e-179
patp:AA96304 Human IGFAM-16 immunoglobulin - <i>Homo sapiens</i>	1254	3.6e-178

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1122 of 1133 bases (99%) identical to a gb:GENBANK-ID:AF067420|acc:AF067420.1 mRNA from *Homo sapiens* (SNC73 protein (SNC73) mRNA). The full amino acid sequence of the protein of the invention was found to have 244 of 253 amino acid residues (96%) identical to, and 247 of 253 amino acid residues (97%) similar to, the 384 amino acid residue ptnr:SPTREMBL-ACC:Q9UP60 protein from *Homo sapiens* (Human) (SNC73 PROTEIN)(Fig. 3B). In addition to smaller changes, the sequence of this invention lacks 46 internal amino acids, when compared to ptnr:SPTREMBL-ACC:Q9UP60

protein from *Homo sapiens* (Human) (SNC73 PROTEIN). NOV99 also has homology to the other proteins shown in the BLASTP data in Table 99D.

Table 99D. NOV99 BLASTP results					
Gene Index / Identifier	Protein / Organism	Length (aa)	Identity (%)	Positive (%)	Expect
gi 229537 prf 752400A	IgA H [<i>Homo sapiens</i>]	475	292/384 (76)	325/384 (84)	e-168
gi 229585 prf 763134A	IgA1 Bur [<i>Homo sapiens</i>]	686	298/383 (77)	323/383 (83)	e-165
gi 223099 prf 0506249A	IgA-alpha1 Bur [<i>Homo sapiens</i>]	472	294/383 (76)	323/383 (83)	e-165
gi 3201900 gb AAC19365.1 (AF067420)	SNC73 protein [<i>Homo sapiens</i>]	384	333/384 (86)	336/384 (86)	e-155
gi 14042015 dbj BAB55072.1 (AK027379)	unnamed protein product [<i>Homo sapiens</i>]	494	321/387 (82)	325/387 (83)	e-150

- 5 This BLASTP data is displayed graphically in the ClustalW in Table 99E. A multiple sequence alignment is given, with the NOV99 protein being shown on line 1 in a ClustalW analysis comparing the protein of the invention with the related protein sequences shown in Table 99D.

Table 99E. ClustalW Alignment of NOV99	
NOV99	(SEQ ID NO:322)
gi 229537	(SEQ ID NO:794)
gi 229585	(SEQ ID NO:795)
gi 223099	(SEQ ID NO:796)
gi 3201900	(SEQ ID NO:797)
gi 14042015	(SEQ ID NO:798)
NOV9910.....20.....30.....40.....50
gi 229537	-----
gi 229585	ESALTZPRSVSGSPGHSVTISCIGTSSNVGDYKYVSWYZZHPGKAPKLI
gi 223099	-----
gi 3201900	-----
gi 14042015	-----
NOV9960.....70.....80.....90.....100
gi 229537	-----
gi 229585	YEVSSRPSGVPDRFSGSKSGBTASLTISGLQAEDEABYYCCSYIGSYVFG
gi 223099	-----
gi 3201900	-----
gi 14042015	-----
NOV99110.....120.....130.....140.....150
NOV99	-----

gi|229537|-----
gi|229585|TGTKVLVIGZPKANPTVTLFPPSSZZLZABKATLVCLISBFYPGAVTVAW
gi|223099|-----
gi|3201900|-----
gi|14042015|-----

160 170 180 190 200
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|-----
gi|229585|KADGSPVKAGVZTTKPSKQSBKYAASSYLSLTPZZWKSHRSYSCQVTHZ
gi|223099|-----
gi|3201900|-----
gi|14042015|-----MELGLR-----WVFLVAFLL

210 220 230 240 250
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|-----EVQLVQSGGGLVVKPCGSLRLSCVASGFSFRDEYMSW
gi|229585|GSTVZKTVAPTZCS EVQLVESGGGVVQAGTSLRLSCTASAFNLSDYAMHW
gi|223099|-----EVQLVESGGGVVQAGTSLRLSCTASAFNLSDYAMHW
gi|3201900|-----
gi|14042015|EGVQ-----CEVQLVESGGGLVVKPCGSLRLSCAASGLSFSTYAMNW

260 270 280 290 300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|IRZTPGKGLZNVSYIGGSGSTLYYADSVKGRFTISRDNAAQKSLYLZMBSL
gi|229585|VRQAPGKGLZNVVALISYGGSBTYIYADSVRGRFTISRBIISKBTLYLZMKTL
gi|223099|VRQAPGKGLZNVVALISYGGSBTYIYADSVRGRFTISRBIISKBTLYLZMKTL
gi|3201900|-----
gi|14042015|VRQAPGKGLZNVSYISRSRDYIYVRDSVKGRFTISRDNAAKNSLYLQMNSL

310 320 330 340 350
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|MAVYYCVSGGIYDSSGP-FDYWGQGLTVTVSSASPTSPKVFPPLSLC
gi|229585|RTZETAVYYCAATBBFBWSTFSLBYWCGELVTVSSASPTSPKVFPPLSLC
gi|223099|RTEDTAVYYCAKLIAVAGTRB---FWGQGLTVTVSLASPTSPKVFPPLSLC
gi|3201900|RTEDTAVYYCAKLIAVAGTRB---FWGQGLTVTVSLASPTSPKVFPPLSLC
gi|14042015|RVDDTAVYYCARDSCNGAICYGFSPWGQGLTVTVSSASPTSPKVFPPLSLC

360 370 380 390 400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|STOPDGNVVIACLVQGGFFPQEPPLSVTWSESGOGVTARNFPSPQDASGDLY
gi|229585|STZPEGBVVIACLVQGGFFPQEPPLSVTWSSZSGZGVTARBFPPSZBASGBLY
gi|223099|STZPEGBVVIACLVQGGFFPQEPPLSVTWSESGZGVTARBFPPSZBASGBLY
gi|3201900|STZPEGBVVIACLVQGGFFPQEPPLSVTWSESGZGVTARBFPPSZBASGBLY
gi|14042015|STOPDGNVVIACLVQGGFFPQEPPLSVTWSESGOGVTARNFPSPQDASGDLY

410 420 430 440 450
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|TTSSQLTLPATQCLAGKSVTCHVKHYTNPSQDVTVPVCPVPSTPPTPSPST
gi|229585|TTSSQLTLPATZCLAGKSVTCHVKHYTTPSZBVTVPVCPVPSTPPTPSPST
gi|223099|TTSSQLTLPATZCLAGKSVTCHVKHYTNPSQDVTVPVCPVPSTPPTPSPST
gi|3201900|TTSSQLTLPATQCLAGKSVTCHVKHYTNPSQDVTVPVCPVPSTPPTPSPST
gi|14042015|TTSSQLTLPATQCLAGKSVTCHVKHYTNPSQDVTVPVCPVPSTPPTPSPST

460 470 480 490 500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

NOV99
gi|229537|PPTPSPSCCHPRLSLHRPALZEDLLLGSEANLTCTLTGLRDASGVFTWTF
gi|229585|PPTPSPSCCHPRLSLHRPALZBLLLGSAZBLTCTLTGLRDASGVFTWTF
gi|223099|PPTPSPSCCHPRLSLHRPALZBLLLGSAZANLTCTLTGLRDASGVFTWTF
gi|3201900|PPTPSPSCCHPRLSLHRPALZEDLLLGSEANLTCTLTGLRDASGVFTWTF
gi|14042015|PPTPSPSCCHPRLSLHRPALZEDLLLGSEANLTCTLTGLRDASGVFTWTF

	510	520	530	540	550
NOV99	SSGKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WNHGEFTFTCTAAHPELKT
gi 229537	SSGKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WDHGTFTCTAAYPESKTP
gi 229585	TSKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WBHGTFTCTAAYPESKTP
gi 223099	TSKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WBHGTFTCTAAYPESKTP
gi 3201900	SSGKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WNHGTFTCTAAYPESKTP
gi 14042015	SSGKSAV	CGPPERDL	CGCYSVSSV	LPGCAEP	WNHGTFTCTAAYPESKTP
	560	570	580	590	600
NOV99	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
gi 229537	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
gi 229585	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
gi 223099	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
gi 3201900	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
gi 14042015	LTATLSKSGNTFRPEV	RLPPPSQ	LALNCL	LVTLTCL	ARGFSPKDVLRW
	610	620	630	640	650
NOV99	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
gi 229537	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
gi 229585	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
gi 223099	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
gi 3201900	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
gi 14042015	LQGSQELPREKYLTWASRQ	EPSQGT	TTFAVTSILRVA	EDWKKGDTFSCM	
	660	670	680		
NOV99	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			
gi 229537	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			
gi 229585	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			
gi 223099	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			
gi 3201900	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			
gi 14042015	VGHEALPLAFTQKTIDRLAGKPTHVNVSVVM	AEVDGTCY			

Table 99F lists the domain description from DOMAIN analysis results against NOV99. This indicates that the NOV99 sequence has properties similar to those of other proteins known to contain this domain.

5

Table 99F. Domain Analysis of NOV99				
gnl Smart smart00407, IGc1, Immunoglobulin C-Type SEQ ID NO:889				
CD-Length = 75 residues, 96.0% aligned				
Score = 46.6 bits (109), Expect = 2e-06				
NOV99:	53	VVIACL	VQGF	PQEP
Sbjct:	2	ATLVCL	VTGF	YP
NOV99:	111	GKSV	TCHV	KHYTNP
Sbjct:	60	GDVY	TCQV	THEGLT

SNC73 was identified by subtractive hybridization between normal mucosa and colorectal cancer tissue as a gene which is down-regulated in colorectal cancer. It is highly homologous to the constant region of immunoglobulin alpha-1 chain. In higher vertebrates there are five classes of antibodies, IgA, IgD, IgE, IgG, and IgM, each with its own class of

10

heavy chain - alpha, delta, epsilon, gamma, and mu, respectively. IgA molecules have alpha chains, IgG molecules have gamma chains, and so on. In addition, there are a number of subclasses of IgG and IgA immunoglobulins; for example, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4) having gamma1, gamma2, gamma3, and gamma4 heavy chains, respectively.

The various heavy chains impart a distinctive conformation to the hinge and tail regions of antibodies and give each class (and subclass) characteristic properties of its own. IgA is the principal class of antibody in secretions (saliva, tears, milk, and respiratory and intestinal secretions). It is transported through secretory epithelial cells from the extracellular fluid into the secreted fluid by the Poly Ig receptor, another type of Fc receptor that is unique to secretory epithelia. IgA serves both to defend against local infection and to prevent access of foreign antigens to the general immunologic system. This function is in accord with the potential role of SNC73 in colorectal cancer.

NOV99 is predicted to be expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, public EST sources, literature sources, and/or RACE sources. Further expression data for NOV99 is provided in Example 2.

The NOV99 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathological disorders described further herein, for example, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, hemophilia, hypercoagulation, immunodeficiencies, graft versus host disease as well as other diseases, disorders and conditions. NOV99 nucleic acids encoding the SNC73-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

The novel nucleic acid of the invention encoding a SNC73-like protein includes the nucleic acid whose sequence is provided in Table 99A, or a fragment thereof. The invention also includes a mutant or variant nucleic acid any of whose bases may be changed from the corresponding base shown in Table 99A while still encoding a protein that maintains its SNC73-like activities and physiological functions, or a fragment of such a nucleic acid. The invention further includes nucleic acids whose sequences are complementary to the sequence of Table 99A, including nucleic acid fragments that are complementary to any of the nucleic acids just described.

The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications. Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject. In the mutant or variant nucleic acids, and their complements, up to about 1% of the bases may be so changed.

The novel protein of the invention includes the SNC73-like protein whose sequence is provided in Table 99B. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in Table 99B while still encoding a protein that maintains its SNC73-like activities and physiological functions, or a functional fragment thereof. In the mutant or variant protein, up to about 4% of the amino acid residues may be so changed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-NOVX Antibodies" section below.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX polypeptides or biologically active portions thereof. Also included in the invention are

nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product “mature” form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term “probes”, as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and

much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

5 The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1
10 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically
15 synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a
20 portion of the nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN
25 MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore,
30 oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a

genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue.

Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the

5 invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a
10 nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or a portion of this nucleotide sequence (e.g., a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162
15 is one that is sufficiently complementary to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base
20 pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or
25 compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific
30 hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs

are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

An NOVX polypeptide is encoded by the open reading frame ("ORF") of an NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated

into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, e.g. from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162; or an anti-sense strand nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162; or of a naturally occurring mutant of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an NOVX protein, such as by measuring a level of an NOVX-encoding nucleic acid in a sample of cells from a subject e.g., detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOVX" can be prepared by isolating a portion SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, that encodes a polypeptide having an NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:2n-1, wherein n is an integer
5 between 1 and 162. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 162.

In addition to the human NOVX nucleotide sequences shown in SEQ ID NOS:2n-1,
10 wherein n is an integer between 1 and 162, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to
15 nucleic acid molecules comprising an open reading frame (ORF) encoding an NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX
20 polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX
25 cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the
30 nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to

describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic

acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, thereby leading to changes in the amino acid sequences of the encoded NOVX proteins, without altering the functional ability of said NOVX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 162. A "non-essential" amino acid residue is a residue that can be altered from the wild-type

sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences SEQ ID NOS:2n, wherein n is an integer between 1 and 162. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 162; more preferably at least about 70% homologous SEQ ID NOS:2n, wherein n is an integer between 1 and 162; still more preferably at least about 80% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 162; even more preferably at least about 90% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 162; and most preferably at least about 95% homologous to SEQ ID NOS:2n, wherein n is an integer between 1 and 162.

An isolated nucleic acid molecule encoding an NOVX protein homologous to the protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 162 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine,

tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved “strong” residues or fully conserved “weak” residues. The “strong” group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the “weak” group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and an NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides

or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an NOVX protein of SEQ ID NOS:2n, wherein n is an integer between 1 and 162, or antisense nucleic acids complementary to an NOVX nucleic acid sequence of SEQ ID NOS:2n-1, wherein n is an integer between 1
5 and 162, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid
10 molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein,
15 antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation
20 start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified
25 nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine,
30 xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil,

beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other. *See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15: 6625-6641.* The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g., Inoue, et al. 1987. Nucl. Acids Res. 15: 6131-6148*) or a chimeric RNA-DNA analogue (*See, e.g., Inoue, et al., 1987. FEBS Lett. 215: 327-330.*

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for an NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of an NOVX cDNA disclosed herein (*i.e.*, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using

standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See*, Hyrup, *et al.*, 1996. *supra*); or as probes or primers for DNA sequence and hybridization (*See*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, Cleavage signal-1 protein H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, *et al.*, 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*, Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In

addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g., Krol, et al., 1988. BioTechniques 6:958-976*) or intercalating agents (*see, e.g., Zon, 1988. Pharm. Res. 5: 539-549*). To this end, the oligonucleotide may be conjugated to another molecule, *e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.*

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in SEQ ID NOS:2n, wherein n is an integer between 1 and 162. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2n, wherein n is an integer between 1 and 162 while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, an NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In

5

10

20

25

30

or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2n, wherein n is an integer between 1 and 162, and retains the functional activity of the NOVX proteins of SEQ ID NOS:2n, wherein
5 n is an integer between 1 and 162.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced
10 in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that
15 position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and
20 Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:2n-1,
25 wherein n is an integer between 1 and 162.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of
30 positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a

polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

5

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, an NOVX "chimeric protein" or "fusion protein" comprises an NOVX polypeptide operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having
10 an amino acid sequence corresponding to an NOVX protein SEQ ID NOS:2n, wherein n is an integer between 1 and 162, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within an NOVX fusion protein the NOVX polypeptide can
15 correspond to all or a portion of an NOVX protein. In one embodiment, an NOVX fusion protein comprises at least one biologically-active portion of an NOVX protein. In another embodiment, an NOVX fusion protein comprises at least two biologically-active portions of an NOVX protein. In yet another embodiment, an NOVX fusion protein comprises at least three biologically-active portions of an NOVX protein. Within the fusion protein, the term
20 "operatively-linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase)
25 sequences. Such fusion proteins can facilitate the purification of recombinant NOVX polypeptides.

In another embodiment, the fusion protein is an NOVX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of NOVX can be increased through use of a heterologous
30 signal sequence.

In yet another embodiment, the fusion protein is an NOVX-immunoglobulin fusion protein in which the NOVX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit

an interaction between an NOVX ligand and an NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of an NOVX cognate ligand.

Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with an NOVX ligand.

10 An NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, 15 alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a 20 chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

25

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific

biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

5 Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene
10 library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce
15 libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate
20 oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

25 In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of an NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per
30 molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can

be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. See, e.g., Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, et al., 1993. *Protein Engineering* 6:327-331.

NOVX Antibodies

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NOs: 2n, wherein n is an

integer between 1 and 162, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of

such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell [Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103]. Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies [Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63].

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin.

Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire sequence of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This

approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al, (Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a

nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

5 In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

10 **F_{ab} Fragments and Single Chain Antibodies**

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective
15 identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab')₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab')₂} fragment; (iii) an F_{ab}
20 fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that
25 have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two
30 immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct

bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another

bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

5 Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using
10 known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

15

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain
20 disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., *J. Exp Med.*, 176: 1191-1195 (1992) and Shopes, *J. Immunol.*, 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as
25 described in Wolff et al. *Cancer Research*, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., *Anti-Cancer Drug Design*, 3: 219-230 (1989).

Immunoconjugates

30 The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have
35 been described above. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such as streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody

of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

5

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic
10 methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds (see below).

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such
15 an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a
20 clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline
25 phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include
30 luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume of the subject to which it is administered. Common ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed.

leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

5 ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or F_{(ab)2}) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*,

tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including
5 fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression
10 vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with
15 vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the
20 recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors
25 include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc
30 (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant

protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g.,
5 Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSecI (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:
10 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the
15 pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are
20 often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1:
25 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477),
30

pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.*, Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of

transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:2n-1, wherein n is an integer between 1 and

162), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the

5

10

15

20

compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, 5 *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or 10 methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of 15 glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, 20 suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, 25 propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic 30 acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

5 In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be
10 obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

15 It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The
20 specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as
25 gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery
30 vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries,

while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, e.g., Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, e.g., nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (e.g., Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an NOVX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the

assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein
 5 determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof,
 10 on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with an NOVX target
 15 molecule. As used herein, a "target molecule" is a molecule with which an NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An NOVX target molecule can be a non-NOVX molecule or an
 20 NOVX protein or polypeptide of the invention. In one embodiment, an NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the
 25 association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with an NOVX target molecule can be accomplished by determining the activity of the
 30 target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*,

luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to an NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate an NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an NOVX protein, wherein determining the ability of the test compound to interact with an NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of an NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent

such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®],
5 Isotridecypoly(ethylene glycol ether)_n, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of
10 complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a
15 fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is
20 incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or
25 activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS
30 (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in

the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to

interact, *in vivo*, forming an NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences, SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome. The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, see, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding

regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

5 Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a
10 noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

15 The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*,
20 blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders,
25 and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in an NOVX gene can be assayed in a
30 biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or

prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NOS:2n-1, wherein n is an integer between 1 and 162, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA,

protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a

disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from an NOVX gene; (ii) an addition of one or more nucleotides to an NOVX gene; (iii) a substitution of one or more nucleotides of an NOVX gene, (iv) a chromosomal rearrangement of an NOVX gene; (v) an alteration in the level of a messenger RNA transcript of an NOVX gene, (vi) aberrant modification of an NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an NOVX gene, (viii) a non-wild-type level of an NOVX protein, (ix) allelic loss of an NOVX gene, and (x) inappropriate post-translational modification of an NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an NOVX gene. A

preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682).

This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see*, Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Q β Replicase (*see*, Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

25 In an alternative embodiment, mutations in an NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent 30 No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing

hundreds or thousands of oligonucleotides probes. *See, e.g., Cronin, et al., 1996. Human Mutation* 7: 244-255; Kozal, *et al., 1996. Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra*. Briefly, a first hybridization array of probes
5 can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation
10 array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of the sample NOVX with the corresponding wild-type (control) sequence.
15 Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g., Naeve, et al., 1995. Biotechniques* 19: 448), including sequencing by mass spectrometry (*see, e.g., PCT*
20 *International Publication No. WO 94/16101; Cohen, et al., 1996. Adv. Chromatography* 36: 127-162; and Griffin, *et al., 1993. Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g., Myers, et al., 1985. Science* 230: 1242. In general, the
25 art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance,
30 RNA/DNA duplexes can be treated with Cleavage signal-1 protein and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing

polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an NOVX sequence, *e.g.*, a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a

denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.*

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer
5 extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230.* Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the
10 oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the
15 molecule (so that amplification depends on differential hybridization; *see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448*) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g., Prossner, 1993. Tibtech. 11: 238*). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,*
20 *Gasparini, et al., 1992. Mol. Cell Probes 6: 1.* It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189.* In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

25 The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an NOVX gene.

30 Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

[illegible]

5
10
15
20

25

30

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic

polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome NEUROPEPTIDE Y/PEPTIDE YY RECEPTOR enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard
5 and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite
10 frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified
15 to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding
20 drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For
30 example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or

downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

[illegible]

5
10

15

15

20

25

30

utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an
5 aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in*
10 *situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the
15 subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is
20 prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, an NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein
30 activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an NOVX protein, a peptide, an NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX

that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering an NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder,

immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

Examples

Example 1. Identification of NOVX clones

The novel NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. Table 100A shows the sequences of the PCR primers used for obtaining different clones. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based

on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain - whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. Table 100B shows a list of these bacterial clones. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

Table 100A. PCR Primers for Exon Linking

NOVX Clone	Primer 1 (5' - 3')	SEQ ID NO	Primer 2 (5' - 3')	SEQ ID NO
NOV1d	TGACTATGGCCTGGAGTTTCCGT	890	TTATTCCAGAGATCCTTGGCAGAAGC	891
NOV3c	ACCCTGGCCCGAGCTCTGAGT	892	GAGCTCAGACGTACTCTCGGGCAG	893
NOV4c	ACCTGTGCAATGGCTTTAATCTTTAG	894	ATCTCTGGCCTTGTCTGGAGTCTTAC	895
NOV5a	GTTTCCCCACCCCGCAGA	896	GAGAGGATCAAAGAACCAGACAGGAG	897
NOV5b	ACGGCGATGACCCCCCAG	898	GGGGGAGGGGCTCAAACAAGA	899
NOV5c	ACGGCGATGACCCCCCAG	900	GGGGGAGGGGCTCAAACAAGA	901
NOV5d	ACGGCGATGACCCCCCAG	902	GGGGGAGGGGCTCAAACAAGA	903
NOV5e	ACGGCGATGACCCCCCAG	904	GGGGGAGGGGCTCAAACAAGA	905
NOV6	GTCGAACCGGGGACCTG	906	TGGGAAGGTGCTCAGCCC	907
NOV7a	CAGCCAAACCCACCTCACCAT	908	TTGGCTGGCTTATATAGCGAGCTCCT	909
NOV7b	CATTGCCAATTCTAATCCATCATTTG	910	TCTTCTAAAGCACAAATAACACCTCCA	911
NOV7c	ATGGCCGAGGCCGCGGAG	912	AGCTTCATTTCATTCTTTTGCAACATCTG	913
NOV8	GGGGACCATGGGGAGCGATC	914	CGAGAGGTTTCTAGGCAGATTTGGAGC	915
NOV15b	ATGGCAGCAGAAAACCATTTCTTTT	916	TCTGTTTATAAGAATGTTCTTTTCCCTAGC	917
NOV18b	AGATGGATGGAACCAATGGCAG	918	ACTGGCCCTGGATGGAAGTGA	919
NOV22c	CACTCCAACAGTTAATGCTTCCCTTG	920	GATGGTCTCGATCTCCTGACCTCTT	921
NOV24b	ATGGCCGAGGCCGCGGAG	922	AGCTTCATTTCATTCTTTTGCAACATCTG	923
NOV26b	GTTCTGTCTGTCTGGACTTTTTCTGT	924	TTTTTGGAGAAAGCTGCAAAAGTTTAT	925
NOV27b	CTCAGTCCCTCGGGCTCATACCTA	926	GGACTCAGAGCTCCTGCCTTTCTGT	927
NOV27c	CTCAGTCCCTCGGGCTCATACCTA	928	CTCAGAGCTCCTGCCTTTCTGTCC	929
NOV29b	GTAGCCACAAGACCGGGTCCG	930	CCCTGGCCTCTTGGAAGTCTTGAT	931
NOV32	CTTGGAGGCTGCAGGTCCTGGAC	932	AGGAGCATCCTTCATCCCACAGAGGT	933
NOV33	GCTGGATTGGGTGATCTCTCAGAGC	934	CTCTTACTCCTGGCAAGCCTGAG	935
NOV36b	ATGCAAGAAACAATTTTTTTTTTGAGA CG	936	TTTTACCATTACCAAGTTTAAGGACG	937
NOV38	CAATGACCTCCTCATTGCTTCTGG	938	CAAAGCCCCAGGTCCTCTTGCTAG	939
NOV39a	CCATCCGAGGCTCCTGAACC	940	CAAAGCCCCAGGTCCTCTTG	941
NOV39b	GTTCTTGAACCAAGGCCATTAC	942	CCTGGGAACTTCATCTTGGTCTCTT	943
NOV42b	ATGGACGGTGAGGCAGTCC	944	CCACACCCTGGCCCATG	945
NOV42c	ATGGACGGTGAGGCAGTCC	946	CCACACCCTGGCCCATG	947
NOV42d	CTGGAGGATGAAGGAAGCAGAGATG	948	CCAGAGAACAAGCAGAGCTCAGAGG	949
NOV46b	GTTTCTGAGCATGGATCCAACCA	950	AAAGGGCAGAGGCTCTTCTCTCAC	951
NOV48b	GCTACCCTCTCTGCTGGCTACCTAAC	952	TTGATTTTACCACCTCCATTGTCTCT	953
NOV50b	AAGATGTCGAGCCCAAGAAAAG	954	TGCTTTGGGAGGTAGCTGGGA	955

NOV51	CTGGCCTAATGAATGTCTCTGAGC	956	TGATCATGGAGGAATAATCTAATATGCCTTA G	957
NOV52	TCCCTGATGTCCAGCTCTGGCT	958	ATAGACTAACTGCACCCACAGGCCTCT	959
NOV56b	ATGGCGAAGATTGAGAAAAACGCT	960	CTCTCAGATCTCCAGGCAGAGTTCAG	961
NOV58b	AGTCTTGCCTTCTTTTGAGCCTAAGTC	962	CACATTCAACATATCTGAGGCTGTGG	963
NOV60a	CGAATTGGCTTCCGAGTAAAAATC	964	TTATTTAAAGGTCAAGGCCTCGAAGTG	965
NOV61	CAGCTGTGCCCTCATCCTTGTGCCTGCT ACGTCC	966	CTCGAGCTTGGTCACTGTGATTCCACCGTG ATATGGTCTGCC	967
NOV63	GGATCCAGCTACCCGATCTGGTGGCTGA CGGGCAGC	968	GTCTGACGCCCTTGCGAGGTGTAGACCTCCTCA CG	969
NOV68	GGCAGCTCCCTCTGGCT	970	TTCCTGTGGGGGCTGG	971
NOV69b	ATGGAAAAAGCATTGAAAATTGACACA	972	TTAAACAGCATAGTTAACCCTAAGTCAGTA GTG	973
NOV70a	TATGCTGTCTATGCTGAATCTTCTAATC TTGTCT	974	TTTAAAGATGTTGAATATTGGCCCCCAC	975
NOV76b	ACCATTACATCATCGTGGCAAATTAA	976	GAAGTCACAAGTGTCTTTCTTCTCAGGA	977
NOV81b	GTCATGCGCTGCCCAAGT	978	CCAATGAGAGTCAGCACTGGAGC	979
NOV87b	TCTCTCATGGCCCCAAAGAC	980	AGTCAGTGCAGCGGGAAGA	981
NOV96a	GGTGCTCTCAGCGTTCTTCCAGTC	982	CTAGTGCTTCTGTACAAAGTCTGGG	983
NOV96d	AGCTGGATTGACAACTTTGTAATGGAG	984	CTCAGTCGTGCTGCTAGTAGGGGT	985
NOV96e	TGAAGCTCACCAGGAGGAAGAAG	986	CTCAGTCGTGCTGCTAGTAGGGGT	987
NOV96f	TCCCATGACCTGCCACTTCC	988	CGCTACCTGCAGCCGCA	989
NOV98	CCGGCCCCGTGTGTGGCA	990	GGGGCTCTGGTGCCAGTCTATG	991
NOV99	CAGCCAAACCCACCTCCACCAT	992	TTTGGCTGGCTTATATAGCGAGCTCCT	993

Physical clone: Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

Table 100B. Physical Clones for PCR products

NOVX Clone	Bacterial Clone
NOV1a	Genomic clone: AC080137
NOV1b	Genomic clone: GMAC036188
NOV1d	Physical clone: GMAC080137_A.698589.M14
NOV2	Physical clone: AC023194
NOV3a	Genomic clone: AL136383
NOV3c	Physical clone: GMAL136383_A.698589.C6
NOV4a	Genomic clone: AC046164
NOV4c	Physical clone: GMAC046164_A.698589.A2
NOV5c	Physical clone: 155643499
NOV5d	Physical clone: 155643499
NOV5e	Physical clone: 127939::153729589.698590.P9
NOV7a	Physical clone: 153634912
NOV7b	Physical clone: 127994::138896306.698587.E16
NOV7c	Physical clone: 127557::CG564550-01.698587.F19
NOV9a	Physical clone: 152568459, AC005961.1, AC068256.2, AI308124, AI307658
NOV10	Physical clone: 126218::CG55964-01.698509.G10
NOV13b	Physical clone: 126388::CG56021-01.698539.G15
NOV15b	Physical clone: 126694::CG56065-01.698561.C17
NOV16a	Physical clone: 126696::CG56067-01.698561.A11
NOV16b	Physical clone: 128769::GMAC084434_A.698655.G14
NOV17b	Physical clone: 128198::GMba430i15_A.698589.F10
NOV17d	Physical clone: 105889::sggc draft ba430i15 20000823.698368.G10
NOV18b	Physical clone: 128205::GMba430i15_D.698590.E4

NOV19b	Physical clone: 128207::GMba430i15 E.698590.C4
NOV21a	Physical clone: 153485226, 145329583
NOV21b	Physical clone: 50119::86375666.244185.F7
NOV22a	Genomic clone: AC011522.6
NOV22c	Physical clone: 161710747.698893.K3
NOV23a	Physical clone: 140117553 146712128 146712112 148412737
NOV23c	Physical clone: 140117553
NOV24b	Physical clone: 127557::CG564550-01.698591.N11
NOV25	Physical clone: 151537975 and 128978463
NOV26a	Physical clone: 127998335
NOV26b	Physical clone: 127561::CG56461-01.698589.G7
NOV27a	Physical clone: 111787393, HSAJ9617
NOV27b	Physical clone: 111787393 EXT.698587.K18
NOV27c	Physical clone: 167695055 170842341 170842333
NOV28	Physical clone: AC004832
NOV29a	Physical clone: AC004832
NOV29b	Physical clone: 112824::COR100399281.698230.B12
NOV29c	Physical clone: AC004832
NOV30	Genomic clone: AC004832
NOV31	Genomic clone: 94329210
NOV33	Physical clone: 153778095, 138978176, 146713055, 105100551, 153777948
NOV34	Physical clone: 125858::GMAC026083 E.698508.J15
NOV35	Physical clone: 114740::AC011711.698329.I10
NOV36a	Physical clone: 152568436, AL132780
NOV36b	Physical clone: 152568436 134511756
NOV37	Physical clone: 151557368, 138195002, 152762569, 152768078
NOV38	Physical clone: 107207::AC061707.698315.F14
NOV39a	Physical clone: 127119::AC061707.698564.H20
NOV39b	Physical clone: 128110::ADENOSINE A3 RECEPTOR.698657.H22.
NOV40	Physical clone: AC068471, AC068471, AV655524, T67857
NOV41a	Physical clone: AC007278, AW242630.1 xn01f05.x1
NOV42a	Physical clone: AC007395, 153103275, 153103263, 153103260
NOV42b	Physical clone: 153103275 153103263 153103260 153103539 153103266 152189065
NOV42c	Physical clone: 54701683 EXT.698433.J23
NOV42d	Physical clone: AC007395 A.698587.M17
NOV43	Genomic clone: AC021773
NOV44	Physical clone: AC023654, 78743598
NOV45	Genomic clone: AC023078
NOV46a	Genomic clone: AC023078
NOV46b	Physical clone: 128292::AC023078 A.698657.G13
NOV46c	Genomic clone: AC023654
NOV46d	Physical clone: 151667972
NOV47	Genomic clone: AF152363
NOV48a	Genomic clone: AC012510.5
NOV48b	Physical clone: 128669::AC012510_5_final.698656.J20
NOV49	Genomic clone: AC011492
NOV50a	Physical clone: 153778754, 122656699
NOV50b	Physical clone: AK001421 A.698657.J10
NOV51	Physical clone: 126131::CG55922-01.698509.O9
NOV52	Physical clone: 153623113, 152186811, 148441423, 148441418, 152186815, 152209564, 126066491, 129293170, 126630256, 124459512
NOV53	Physical clone: 151222559
NOV54a	Physical clone: 153512063
NOV55	Genomic clone: AL138816.12, AL158192.12
NOV56a	Genomic clone: AC019100.4
NOV57	Physical clone: 50222151, 150222148
NOV54b	Physical clone: 164698940
NOV56b	Physical clone: 55048::nh0443k08 A.698002.M9

Example 2. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection System.

5 Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on
10 metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoimmune diseases), Panel CNSD.01 (containing central nervous system samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of
15 agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

20 First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β -actin and GAPDH). Normalized RNA (5 μ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

25 In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random hexamers according to the manufacturer's instructions. Reactions containing up to 10 μ g of total RNA were performed in a volume of 20 μ l and incubated for 60 minutes at 42°C. This reaction can be scaled up to 50 μ g of total RNA in a final volume of 100 μ l. sscDNA samples
30 are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a

similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60°C, primer optimal T_m = 59°C, maximum primer difference = 2°C, probe does not have 5'G, probe T_m must be 10°C greater than primer T_m , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95°C 10 min, then 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS

cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

General_screening_panel_v1.4

The plates for Panel 1.4 include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panel 1.4 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panel 1.4 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on Panel 1.4 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult

lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

5 **Panels 2D and 2.2**

 The plates for Panels 2D and 2.2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from
10 human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI or CHTN). This analysis provides a gross
15 histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from
20 autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

Panel 3D

 The plates of Panel 3D are comprised of 94 cDNA samples and two control samples.
25 Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic
30 cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2μg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^6 cells/ml in DMEM 5% FCS (Hyclone), 100μM non essential amino acids (Gibco), 1mM

sodium pyruvate (Gibco), mercaptoethanol ($5.5 \times 10^{-5} \text{M}$) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μ g/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μ g/ml anti-CD28 (Pharmingen) and 3 μ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol $5.5 \times 10^{-5} \text{M}$ (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS

(Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile
5 dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10⁶cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5 μ g/ml or anti-CD40 (Pharmingen) at approximately 10 μ g/ml and IL-4 at 5-10ng/ml. Cells were harvested for RNA preparation at
10 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μ g/ml anti-CD28 (Pharmingen) and 2 μ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10⁵-10⁶cells/ml in DMEM 5% FCS (Hyclone), 100 μ M
15 non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1 μ g/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1 μ g/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM
20 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were
25 washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

30 The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5x10⁵cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5x10⁵cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μ M non essential amino acids

(Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1µg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco). CCD106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20°C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300µl of RNase-free water and 35µl buffer (Promega) 5µl DTT, 7µl RNAsin and 8µl DNase were added. The tube was incubated at 37°C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80°C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

5 Surgical specimens of diseased colon from patients with ulcerative colitis and Crohn's disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female
10 patients. Four of the patients were taking lebid and two were on phenobarbital.

 Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1-antitrypsin deficiencies.
15 Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

 In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

20 AI = Autoimmunity
 Syn = Synovial
 Normal = No apparent disease
 Rep22 /Rep20 = individual patients
 RA = Rheumatoid arthritis
25 Backus = From Backus Hospital
 OA = Osteoarthritis
 (SS) (BA) (MF) = Individual patients
 Adj = Adjacent tissue
 Match control = adjacent tissues
30 -M = Male
 -F = Female
 COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed, the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

Patient 2	Diabetic Hispanic, overweight, not on insulin
Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
Patient 10	Diabetic Hispanic, overweight, on insulin
Patient 11	Nondiabetic African American and overweight
Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose

Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated

Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups:

kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were

examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

- 5 PSP = Progressive supranuclear palsy
- Sub Nigra = Substantia nigra
- Glob Palladus= Globus palladus
- Temp Pole = Temporal pole
- Cing Gyr = Cingulate gyrus
- 10 BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and
15 Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from
20 Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD
25 senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus;
30 the parietal cortex shows moderate neuronal death in the late stages of the disease; the occipital cortex is spared in AD and therefore acts as a "control" region within AD patients. Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy

Control = Control brains; patient not demented, showing no neuropathology

Control (Path) = Control brains; patient not demented but showing severe AD-like pathology

SupTemporal Ctx = Superior Temporal Cortex

Inf Temporal Ctx = Inferior Temporal Cortex

NOV9a and NOV9b

Expression of gene NOV9a and variant NOV9b was assessed using the primer-probe sets Ag2930, Ag4297 and Ag573, described in Tables AA, AB and AC. Results of the RTQ-PCR runs are shown in Tables AD, AE, AF, AG, and AH. Please note that the probe and primer set Ag4297 do not match the NOV9b variant.

Table AA. Probe Name Ag2930

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-attggattttccaactccatct-3'	22	1213	994
Probe	TET-5'-tcccattgtctatgcatttatgaatga-3'-TAMRA	27	1239	995
Reverse	5'-tgcaataacaaaactgcagacaa-3'	22	1285	996

Table AB. Probe Name Ag4297

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tctacaccaccttcacatctgt-3'	22	1007	997
Probe	TET-5'-ctgcctcttatggagaagaaacgagctg-3'-TAMRA	28	1042	998
Reverse	5'-caccactgtcaccatcataatg-3'	22	1071	999

Table AC. Probe Name Ag573

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgtccagtctaccgctgttgtg-3'	22	738	1000
Probe	TET-5'-agaaatcctcactatgacctgcattgctgtg-3'-TAMRA	31	762	1001
Reverse	5'-cacaagtccttggtgccttt-3'	20	794	1002

Table AD. Panel 1.1

Tissue Name	Rel. Exp.(%)	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)	Rel. Exp.(%)
-------------	--------------	--------------	-------------	--------------	--------------

	Ag573, Run 109566077	Ag573, Run 111156622		Ag573, Run 109566077	Ag573, Run 111156622
Adrenal gland	0.0	0.8	Renal ca. UO-31	0.1	6.0
Bladder	0.0	0.0	Renal ca. RXF 393	0.0	0.0
Brain (amygdala)	0.0	0.6	Liver	0.0	0.0
Brain (cerebellum)	0.1	0.1	Liver (fetal)	0.0	0.0
Brain (hippocampus)	0.0	5.3	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (substantia nigra)	10.7	13.7	Lung	0.0	0.0
Brain (thalamus)	0.1	4.9	Lung (fetal)	0.0	1.5
Cerebral Cortex	9.3	11.2	Lung ca. (non-s.cell) HOP-62	4.5	0.0
Brain (fetal)	0.0	0.8	Lung ca. (large cell) NCI-H460	0.0	0.0
Brain (whole)	1.3	6.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	2.2	0.0
astrocytoma SF-539	0.0	1.3	Lung ca. (non-sm. cell) A549	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Lung ca. (s.cell var.) SHP-77	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (small cell) LX-1	0.0	0.0
glioma U251	0.0	0.0	Lung ca. (small cell) NCI-H69	43.5	47.0
glioma SF-295	1.3	0.0	Lung ca. (squam.) SW 900	0.0	0.0
glioma SNB-19	0.0	0.5	Lung ca. (squam.) NCI-H596	100.0	98.6
glio/astro U87-MG	0.0	1.5	Lymph node	0.0	0.0
neuro*; met SK-N-AS	2.6	0.1	Spleen	0.0	0.0
Mammary gland	0.0	3.8	Thymus	0.0	0.0

Breast ca. BT-549	0.0	0.3	Ovary	0.0	0.6
Breast ca. MDA-N	28.3	39.8	Ovarian ca. IGROV-1	0.0	0.0
Breast ca.* (pl.ef) T47D	0.0	0.0	Ovarian ca. OVCAR-3	0.2	1.8
Breast ca.* (pl.ef) MCF-7	0.3	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0	Ovarian ca. OVCAR-5	0.0	0.1
Small intestine	0.0	0.0	Ovarian ca. OVCAR-8	2.5	16.3
Colorectal	0.0	0.0	Ovarian ca.* (ascites) SK-OV-3	0.8	4.6
Colon ca. HT29	0.0	0.0	Pancreas	2.5	0.0
Colon ca. CaCo-2	0.0	0.0	Pancreatic ca. CAPAN 2	0.0	0.0
Colon ca. HCT-15	0.0	0.0	Pituitary gland	1.5	2.7
Colon ca. HCT-116	0.0	0.0	Placenta	0.0	0.0
Colon ca. HCC-2998	0.0	0.0	Prostate	0.0	0.0
Colon ca. SW480	0.0	0.0	Prostate ca.* (bone met) PC-3	0.1	0.0
Colon ca.* SW620 (SW480 met)	1.9	0.0	Salivary gland	0.1	0.0
Stomach	0.0	0.0	Trachea	0.0	0.0
Gastric ca. (liver met) NCI-N87	0.1	0.0	Spinal cord	0.0	2.4
Heart	3.8	19.6	Testis	0.0	2.0
Skeletal muscle (Fetal)	0.0	0.3	Thyroid	0.0	2.3
Skeletal muscle	0.0	0.0	Uterus	0.0	0.0
Endothelial cells	0.0	0.0	Melanoma M14	0.0	0.5
Heart (Fetal)	11.0	31.0	Melanoma LOX IMVI	0.0	0.0
Kidney	1.3	12.2	Melanoma UACC-62	0.0	0.0
Kidney (fetal)	0.0	3.9	Melanoma SK-MEL-28	0.0	0.4

Renal ca. 786-0	0.0	5.4	Melanoma* (met) SK-MEL-5	0.0	0.0
Renal ca. A498	10.5	16.4	Melanoma Hs688(A).T	0.0	0.2
Renal ca. ACHN	85.3	100.0	Melanoma* (met) Hs688(B).T	0.0	0.0
Renal ca. TK-10	60.7	59.0			

Table AE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2930, Run 158090377	Rel. Exp.(%) Ag2930, Run 165701939	Tissue Name	Rel. Exp.(%) Ag2930, Run 158090377	Rel. Exp.(%) Ag2930, Run 165701939
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	10.1	0.0
Pancreas	0.0	0.0	Renal ca. 786-0	2.3	7.8
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	61.6	44.8
Adrenal gland	0.0	0.0	Renal ca. RXF 393	0.0	0.0
Thyroid	1.6	0.0	Renal ca. ACHN	16.8	100.0
Salivary gland	0.0	0.0	Renal ca. UO-31	10.1	15.0
Pituitary gland	3.1	0.0	Renal ca. TK-10	22.5	52.5
Brain (fetal)	9.7	3.6	Liver	0.0	0.0
Brain (whole)	11.0	24.3	Liver (fetal)	1.3	0.0
Brain (amygdala)	23.2	9.3	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	0.0	0.0	Lung	0.0	0.0
Brain (hippocampus)	100.0	3.4	Lung (fetal)	0.0	15.0
Brain (substantia nigra)	2.7	19.5	Lung ca. (small cell) LX-1	0.0	0.0
Brain (thalamus)	23.3	6.8	Lung ca. (small cell) NCI-H69	39.2	16.0
Cerebral Cortex	33.4	0.0	Lung ca. (s.cell var.) SHP-77	0.0	0.0

Spinal cord	6.4	7.0	Lung ca. (large cell)NCI-H460	0.0	0.0
glio/astro U87-MG	2.6	4.9	Lung ca. (non-sm. cell) A549	0.0	0.0
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.0	0.0
neuro*; met SK-N-AS	4.2	0.0	Lung ca. (non-s.cl) NCI-H522	0.0	0.0
astrocytoma SF-539	5.0	0.0	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Lung ca. (squam.) NCI-H596	8.4	6.7
glioma SNB-19	2.1	0.0	Mammary gland	18.3	0.0
glioma U251	0.0	0.0	Breast ca.* (pl.ef) MCF-7	1.4	4.6
glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
Heart (fetal)	18.4	0.0	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	2.4	4.1	Breast ca. BT-549	10.7	5.8
Skeletal muscle (fetal)	1.5	0.0	Breast ca. MDA-N	35.1	4.7
Skeletal muscle	0.0	0.0	Ovary	0.0	0.0
Bone marrow	0.0	0.0	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	1.8	0.0	Ovarian ca. OVCAR-5	0.0	0.0
Lymph node	0.0	11.6	Ovarian ca. OVCAR-8	12.9	4.9
Colorectal	0.0	0.0	Ovarian ca. IGROV-1	0.0	0.0
Stomach	1.3	3.3	Ovarian ca.* (ascites) SK-OV-3	2.9	6.5
Small intestine	1.7	0.0	Uterus	0.0	0.0

Colon ca. SW480	0.0	0.0	Placenta	0.0	0.0
Colon ca.* SW620(SW480 met)	0.0	0.0	Prostate	0.0	0.0
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC- 3	0.0	0.0
Colon ca. HCT- 116	0.0	0.0	Testis	2.4	0.0
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	0.0	0.0	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC- 2998	0.0	0.0	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma M14	0.0	0.0
Bladder	0.0	0.0	Melanoma LOX IMVI	0.0	0.0
Trachea	0.0	0.0	Melanoma* (met) SK- MEL-5	0.0	0.0
Kidney	0.0	8.5	Adipose	1.3	0.0

Table AF. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2930, Run 158090382	Tissue Name	Rel. Exp.(%) Ag2930, Run 158090382
Normal Colon	1.2	Kidney Margin 8120608	0.6
CC Well to Mod Diff (ODO3866)	0.4	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.3	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	3.2
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	1.7
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.0
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.5
CC Margin (ODO3921)	2.2	Thyroid Cancer 064010	0.3

CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.2
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.4
Prostate Cancer (OD04410)	0.4	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Prostate Cancer (OD04720-01)	1.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.2
Normal Lung 061010	0.2	Breast Cancer 9100266	3.5
Lung Met to Muscle (ODO4286)	0.5	Breast Margin 9100265	0.3
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	1.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.4	Normal Liver	0.0
Lung Cancer (OD04404)	1.9	Liver Cancer 064003	0.2
Lung Margin (OD04404)	0.0	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.4
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237- 01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237- 02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.2	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	5.0
Normal Kidney	1.6	Bladder Cancer (OD04718-01)	0.0

SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	2.7	JM1- pre-B-cell lymphoma	0.0
Cerebellum	7.3	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	39.0
NCI-N417- Small cell lung cancer	100.0	Caki-2- Clear cell renal carcinoma	4.6
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	16.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	5.0
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0

KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinoma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinoma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	8.7	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table AH. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2930, Run 158090383	Tissue Name	Rel. Exp.(%) Ag2930, Run 158090383
-------------	--	-------------	--

Secondary Th1 act	5.9	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	9.9
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	15.2
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	22.1
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.2
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	10.7
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0

Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	7.9	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	6.3
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	100.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag2930 Expression of the NOV9A gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

General_screening_panel_v1.5 Summary: Ag4297 Expression of the NOV9A gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

5 **Panel 1.1 Summary:** Ag573 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of the NOV9A gene, a putative neuropeptide Y receptor, in lung cancer and renal cancer cell lines (CTs=23-26). Significant expression is also seen in a cluster of breast cancer cell lines. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel
10 and as a marker to detect the presence of these cancers. Neuropeptide Y, which controls vasoconstriction and feeding behavior, is expressed in breast cancer (see ref. below). Furthermore, peptide receptors in human tumors represent clinically relevant targets for both

Furtinger S, Pirker S, Czech T, Baumgartner C, Ransmayr G, Sperk G. Plasticity of Y1 and Y2 receptors and neuropeptide Y fibers in patients with temporal lobe epilepsy. *J Neurosci* 2001 Aug 1;21(15):5804-12

Marked expression of neuropeptide Y (NPY) and its Y2 receptors in hippocampal mossy fibers has been reported in animal models of epilepsy. Because NPY can suppress glutamate release by activating presynaptic Y2 receptors, these changes have been proposed as an endogenous protective mechanism. Therefore, we investigated whether similar changes in the NPY system may also take place in human epilepsy. We investigated Y1 and Y2 receptor binding and NPY immunoreactivity in hippocampal specimens that were obtained at surgery from patients with temporal lobe epilepsy and in autopsy controls. Significant increases in Y2 receptor binding (by 43-48%) were observed in the dentate hilus, sectors CA1 to CA3, and subiculum of specimens with, but not in those without, hippocampal sclerosis. On the other hand, Y1 receptor binding was significantly reduced (by 62%) in the dentate molecular layer of sclerotic specimens. In the same patients, the total lengths of NPY immunoreactive (NPY-IR) fibers was markedly increased (by 115-958%) in the dentate molecular layer and hilus, in the stratum lucidum of CA3, and throughout sectors CA1 to CA3 and the subiculum, as compared with autopsies. In nonsclerotic specimens, increases in lengths of NPY-IR fibers were more moderate and statistically not significant. NPY mRNA was increased threefold in hilar interneurons of sclerotic and nonsclerotic specimens. It is suggested that abundant sprouting of NPY fibers, concomitant upregulation of Y2 receptors, and downregulation of Y1 receptors in the hippocampus of patients with Ammon's horn sclerosis may be endogenous anticonvulsant mechanisms.

PMID: 11466452

Rahmouni K, Haynes WG. Leptin signaling pathways in the central nervous system: interactions between neuropeptide Y and melanocortins. *Bioessays*. 2001 Dec;23(12):1095-9.

No other hormone has drawn more attention than leptin in recent studies on the control of appetite, body weight and obesity. This hormone is produced by adipose tissue and enters the brain via a saturable specific transport mechanism. Leptin acts in the hypothalamus to modulate food intake and heat production as well as several other neuroendocrine pathways. The mechanisms through which leptin exerts its central nervous effects are now better understood. Proopiomelanocortin- and neuropeptide Y-containing neurons in the hypothalamus have emerged as potent candidate mediators of leptin action. These two neuropeptides have been shown to exert opposing effects using different pathways. Recently, Cowley et al. (2001) described a new circuit in the regulation of neuronal activity by leptin

with an interaction between these two pathways. These data add complexity to the mechanisms by which leptin achieves its effects in the central nervous system, but they also offer potential mechanisms to explain the phenomenon of leptin resistance observed in obesity. Copyright 2001 John Wiley & Sons, Inc.

5 PMID: 11746228

Michalkiewicz M, Michalkiewicz T, Kreulen DL, McDougall SJ.

Increased blood pressure responses in neuropeptide Y transgenic rats. *Am J Physiol Regul Integr Comp Physiol* 2001 Aug;281(2):R417-26

10 Considering the coexistence of neuropeptide Y (NPY) and norepinephrine in perivascular sympathetic nerves and the known vasoconstrictor cooperation of NPY with norepinephrine, we investigated the involvement of NPY in long-term control of cardiovascular functions using NPY transgenic (NPY-tg) rats. These rats were developed by injection of the rat (Sprague-Dawley) pronuclei with a 14.5-kb clone of the rat structural NPY gene. When compared with nontransgenic littermates, NPY concentrations were significantly
15 increased in a number of cardiovascular tissues of NPY-tg hemizygotes. Direct basal mean arterial pressure and heart rate were not changed, but calculated total vascular resistance was significantly increased in NPY-tg subjects. Arterial pressure increases, in response to norepinephrine injection, were greater in the NPY-tg rats. Also, the hypotension and bradycardia in response to hemorrhage were significantly reduced in NPY-tg subjects. These
20 results indicate that NPY, when expressed in increased amounts, potentiates the pressor effects of norepinephrine and contributes to maintaining blood pressure during hemorrhage, but it does not alter resting blood pressure. These transgenic rats will facilitate studies of the role of NPY signaling in cardiovascular regulation, particularly regarding its functional cooperation with norepinephrine.

25 PMID: 11448843

Horackova M, Slavikova J, Byczko Z. Postnatal development of the rat intrinsic cardiac nervous system: a confocal laser scanning microscopy study in whole-mount atria. *Tissue Cell* 2000 Oct;32(5):377-88

30 We used confocal laser scanning microscopy and fluorescent immunohistochemistry to study the developmental pattern and distribution of specific neuronal phenotypes within the intrinsic cardiac nervous system in whole-mount atrial preparations from newborn to 5 week old rats. Individual ganglia and neuronal cell bodies were localized by means of two general neuronal markers: protein gene product 9.5 (PGP) and microtubule-associated protein two (MAP). In rats \leq 2 weeks old there were two main subpopulations of intrinsic neurons

located in the intraatrial septum and around the origin of the superior vena cava. The more abundant was a population of strongly tyrosine hydroxylase (TH) immunoreactive (IR) neurons (10-40 microm in diameter) most of which were also PGP-IR. The second, less numerous (approximately 60-70% than the TH-IR group) type of neurons exhibited ChAT-IR which colocalized with MAP-IR. Towards the end of the second postnatal week and during the third, the ganglia containing these neurons became more numerous and their localization also included tissues around the origins of the inferior vena cava and the pulmonary veins, as well as both atrial walls close to the AV junction. During the second and third postnatal weeks, when the extrinsic innervation of the adrenergic and cholinergic phenotypes largely increases, the intrinsic innervation also changed greatly, and around the 21st postnatal day it appeared to acquire mature characteristics. The TH-IR neurons changed their characteristics and formed two types of ganglia. The larger ganglia containing large cells (20-40 microm in diameter) expressed TH-IR mostly close to their inner body surface (approximately 80-90% of identified neurons). Most of these neurons also expressed neuropeptide Y (NPY)-IR, specifically around their nuclei. The second type of small strongly TH-IR neurons (approximately 10% of all identified neurons) were contained in smaller groups (20-50 cells) which were usually embedded into much larger ganglia (100-400 cells), containing large (20-50 microm) neurons. Unlike all other intrinsic neurons, these small TH-IR cells did not exhibit any PGP-IR or MAP-IR. The number of ChAT-IR neurons increased at this stage, reaching approximately 90% of the neurons identified by the general neuronal markers. These neurons were surrounded by a rich network of cholinergic varicose nerve fibers, some of which were likely of an extrinsic origin. We have also identified relatively small ganglia expressing immunoreactivity to vasoactive intestinal polypeptide (VIP), and to substance P (SP). The presented data indicate that the phenotypes of intrinsic neurons in the rat heart change greatly during the first month of postnatal development. This may be at least partially related to the development and maturation of functional extrinsic nervous control of the heart.

PMID: 11201277

Panel 1.3D Summary: Ag2930 The expression of the NOV9a gene was assessed in two independent runs on this panel. Low but significant levels of expression are seen in kidney cancer cells and a lung cancer cell, consistent with Panel 1.1. Please see the previous panel for discussion of utility of this gene in cancer.

The expression in this panel also confirms expression of this gene product in the brain. Please see Panel 1.1 for discussion of utility of this gene in the central nervous system.

Panel 2D Summary: Ag2930 The NOV9a gene is expressed at low but significant levels in kidney cancer samples in this panel but not in the adjacent normal tissue samples (CTs=30-32). This expression is consistent with results in the preceding panels. This suggests that expression of this gene can be used as a diagnostic marker for the presence of kidney cancer. Furthermore, therapeutic inhibition of the gene product could potentially be used in the treatment of kidney cancer.

Panel 3D Summary: Ag2930 The NOV9a gene expression is restricted to NCI-N417, a small cell lung cancer cell line (CT=33.81). Expression of this gene can therefore be used for the diagnosis and treatment of this cancer.

Panel 4D Summary: Ag2930 Expression of the NOV9a gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

NOV4b

Expression of gene NOV4b was assessed using the primer-probe set Ag2955, described in Table BA. Results of the RTQ-PCR runs are shown in Tables BB, BC and BD.

Table BA. Probe Name Ag2955

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-caatgcaatctgttggactttt-3'	22	33	1003
Probe	TET-5'-tctctctctgggatggattttatccat-3'-TAMRA	26	59	1004
Reverse	5'-gttcttccagtgtggcaaataa-3'	22	94	1005

Table BB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag2955, Run 216860933	Tissue Name	Rel. Exp.(%) Ag2955, Run 216860933
Adipose	8.1	Renal ca. TK-10	8.0
Melanoma* Hs688(A).T	8.5	Bladder	18.9
Melanoma* Hs688(B).T	5.8	Gastric ca. (liver met.) NCI-N87	35.6
Melanoma* M14	1.1	Gastric ca. KATO III	2.5
Melanoma* LOXIMVI	31.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	3.1	Colon ca. SW480	4.4
Squamous cell carcinoma SCC-4	1.3	Colon ca.* (SW480 met) SW620	1.7
Testis Pool	0.0	Colon ca. HT29	3.9

Prostate ca.* (bone met) PC-3	36.9	Colon ca. HCT-116	5.0
Prostate Pool	0.0	Colon ca. CaCo-2	5.9
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	10.5	Colon ca. Colo-205	3.3
Ovarian ca. SK-OV-3	47.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	4.5	Colon Pool	1.6
Ovarian ca. OVCAR-5	6.5	Small Intestine Pool	2.4
Ovarian ca. IGROV-1	21.0	Stomach Pool	4.3
Ovarian ca. OVCAR-8	3.1	Bone Marrow Pool	5.3
Ovary	4.3	Fetal Heart	2.9
Breast ca. MCF-7	0.0	Heart Pool	3.8
Breast ca. MDA-MB-231	29.7	Lymph Node Pool	8.0
Breast ca. BT 549	3.0	Fetal Skeletal Muscle	1.6
Breast ca. T47D	6.3	Skeletal Muscle Pool	7.6
Breast ca. MDA-N	1.6	Spleen Pool	2.4
Breast Pool	0.7	Thymus Pool	1.2
Trachea	2.5	CNS cancer (glio/astro) U87-MG	2.0
Lung	1.5	CNS cancer (glio/astro) U-118-MG	1.2
Fetal Lung	9.6	CNS cancer (neuro;met) SK-N-AS	5.3
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	2.6
Lung ca. LX-1	2.4	CNS cancer (astro) SNB-75	6.7
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	9.3
Lung ca. SHP-77	1.7	CNS cancer (glio) SF-295	62.9
Lung ca. A549	3.0	Brain (Amygdala) Pool	1.3
Lung ca. NCI-H526	0.0	Brain (cerebellum)	4.0
Lung ca. NCI-H23	59.0	Brain (fetal)	7.2
Lung ca. NCI-H460	2.1	Brain (Hippocampus) Pool	2.1
Lung ca. HOP-62	3.0	Cerebral Cortex Pool	3.0

Tissue Name	Rel. Exp.(%) Ag2955, Run 167906363	Tissue Name	Rel. Exp.(%) Ag2955, Run 167906363
Liver adenocarcinoma	8.5	Kidney (fetal)	4.4
Pancreas	0.0	Renal ca. 786-0	3.2
Pancreatic ca. CAPAN 2	64.6	Renal ca. A498	15.1
Adrenal gland	0.0	Renal ca. RXF 393	10.7
Thyroid	1.5	Renal ca. ACHN	6.5
Salivary gland	1.1	Renal ca. UO-31	2.3
Pituitary gland	0.0	Renal ca. TK-10	13.2
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	2.7	Liver (fetal)	0.0
Brain (amygdala)	12.1	Liver ca. (hepatoblast) HepG2	8.5
Brain (cerebellum)	1.0	Lung	6.4
Brain (hippocampus)	0.0	Lung (fetal)	2.9
Brain (substantia nigra)	6.2	Lung ca. (small cell) LX-1	5.5
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	3.9	Lung ca. (s.cell var.) SHP-77	4.1
Spinal cord	1.9	Lung ca. (large cell)NCI-H460	1.1
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	19.2
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	26.1
astrocytoma SW1783	7.6	Lung ca. (non-s.cell) HOP-62	12.3

neuro*; met SK-N-AS	3.5	Lung ca. (non-s.cl) NCI-H522	46.3
astrocytoma SF-539	13.2	Lung ca. (squam.) SW 900	16.7
astrocytoma SNB-75	3.6	Lung ca. (squam.) NCI-H596	2.8
glioma SNB-19	16.5	Mammary gland	0.0
glioma U251	26.6	Breast ca.* (pl.ef) MCF-7	3.0
glioma SF-295	6.0	Breast ca.* (pl.ef) MDA-MB-231	7.1
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	11.5
Heart	1.1	Breast ca. BT-549	1.1
Skeletal muscle (fetal)	3.3	Breast ca. MDA-N	5.7
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	21.2
Thymus	0.0	Ovarian ca. OVCAR-4	7.4
Spleen	0.0	Ovarian ca. OVCAR-5	14.3
Lymph node	0.0	Ovarian ca. OVCAR-8	4.7
Colorectal	1.4	Ovarian ca. IGROV-1	38.4
Stomach	0.9	Ovarian ca.* (ascites) SK-OV-3	100.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	6.8	Placenta	0.0
Colon ca.* SW620(SW480 met)	7.6	Prostate	0.0
Colon ca. HT29	6.1	Prostate ca.* (bone met)PC-3	13.1
Colon ca. HCT-116	4.2	Testis	0.0
Colon ca. CaCo-2	19.3	Melanoma Hs688(A).T	14.7
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	5.4
Colon ca. HCC-2998	3.6	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	20.6	Melanoma M14	0.0
Bladder	13.4	Melanoma LOX IMVI	25.5
Trachea	3.6	Melanoma* (met)	1.7

		SK-MEL-5	
Kidney	5.1	Adipose	13.9

Table BD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2955, Run 164306318	Tissue Name	Rel. Exp.(%) Ag2955, Run 164306318
Secondary Th1 act	0.0	HUVEC IL-1beta	6.6
Secondary Th2 act	0.0	HUVEC IFN gamma	20.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	15.7
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	8.4
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	6.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	8.7
Primary Th2 act	0.0	Microvascular Dermal EC none	7.7
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	6.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	9.5
Primary Th2 rest	0.0	Small airway epithelium none	5.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	19.9
CD45RA CD4 lymphocyte act	13.5	Coronary artery SMC rest	22.4
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	12.0
CD8 lymphocyte act	0.0	Astrocytes rest	37.6
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	29.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	25.7
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	100.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	4.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.3
LAK cells IL-2	0.0	Liver cirrhosis	7.5
LAK cells IL-2+IL-12	0.0	Lupus kidney	5.0

LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	31.9
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	52.1
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	39.5
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	16.6
Two Way MLR 3 day	3.9	NCI-H292 IFN gamma	35.4
Two Way MLR 5 day	0.0	HPAEC none	4.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	3.0
PBMC rest	0.0	Lung fibroblast none	4.9
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	5.5	Lung fibroblast IL-4	9.9
Ramos (B cell) none	0.0	Lung fibroblast IL-9	6.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	1.7
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	11.7
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	59.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	80.7
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	35.4
Dendritic cells none	10.6	Dermal fibroblast IFN gamma	70.2
Dendritic cells LPS	4.5	Dermal fibroblast IL-4	72.7
Dendritic cells anti-CD40	0.0	IBD Colitis 2	4.2
Monocytes rest	7.2	IBD Crohn's	0.0
Monocytes LPS	2.4	Colon	6.7
Macrophages rest	3.3	Lung	22.4
Macrophages LPS	0.0	Thymus	23.2
HUVEC none	7.3	Kidney	0.0
HUVEC starved	19.6		

General_screening_panel_v1.4 Summary: Ag2955 Highest expression of the NOV4b gene is seen in a pancreatic cancer cell line (CT=32.6). Low but significant levels of expression are also seen in melanoma, lung, brain, ovarian, breast and prostate cancer cell lines. Thus, expression of this gene might be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic inhibition of this gene product may be useful in the treatment of melanoma, lung, brain, ovarian, breast and prostate cancers.

5

10

15

20

Table CA. Probe Name Ag2957

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-caggcctctgcactttgac-3'	19	438	1006
Probe	TET-5'-ctgtctcgtgggtatgccaccctggg-3'-TAMRA	25	459	1007
Reverse	5'-ccaaattctgggttgaagaact-3'	22	492	1008

Table CB. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag2957, Run 216861284	Tissue Name	Rel. Exp.(%) Ag2957, Run 216861284
Adipose	1.8	Renal ca. TK-10	6.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0

Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	4.6
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	24.8
Placenta	41.5	Colon cancer tissue	1.4
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	2.7	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	1.7	Fetal Skeletal Muscle	8.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	3.0
Trachea	2.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	2.9	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	24.5	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.9	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0

Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	6.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	1.7
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	4.7
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	31.4	Brain (whole)	0.0
Liver ca. HepG2	7.6	Spinal Cord Pool	22.4
Kidney Pool	1.7	Adrenal Gland	0.0
Fetal Kidney	100.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	2.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table CC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2957, Run 167906390	Tissue Name	Rel. Exp.(%) Ag2957, Run 167906390
Liver adenocarcinoma	0.0	Kidney (fetal)	34.2
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.1	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.2	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	100.0
Brain (whole)	0.0	Liver (fetal)	2.6
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	2.7
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.2	Lung (fetal)	0.8
Brain (substantia nigra)	0.3	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	1.5

Spinal cord	23.8	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.3
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.2	Lung ca. (squam.) SW 900	2.1
astrocytoma SNB-75	0.1	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	1.7
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.6	Breast ca. BT-549	0.3
Skeletal muscle (fetal)	2.2	Breast ca. MDA-N	0.0
Skeletal muscle	1.5	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	2.3	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.1	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.1
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.8
Colon ca.* SW620(SW480 met)	0.3	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	2.2	Melanoma Hs688(A).T	0.0

Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.1
Trachea	0.2	Melanoma* (met) SK-MEL-5	0.0
Kidney	30.6	Adipose	0.3

Table CD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2957, Run 170858345	Tissue Name	Rel. Exp.(%) Ag2957, Run 170858345
Normal Colon	0.0	Kidney Margin 8120608	68.8
CC Well to Mod Diff (ODO3866)	1.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	24.3
CC Gr.2 rectosigmoid (ODO3868)	13.6	Kidney Cancer 9010320	3.4
CC Margin (ODO3868)	1.9	Kidney Margin 9010321	85.9
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	1.6
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.0
CC Margin (ODO3921)	1.2	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	3.3
Lung Margin (OD04451-02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	1.3	Breast Cancer (OD04590-01)	3.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin	2.6	Breast Cancer	0.0

Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	100.0	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	50.3	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	58.6	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	0.0

Table CE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2957, Run 164306319	Tissue Name	Rel. Exp.(%) Ag2957, Run 164306319
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.8	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0

Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	1.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	5.8
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	2.3
Macrophages rest	0.0	Lung	0.0

Macrophages LPS	0.0	Thymus	100.0
HUVEC none	0.0	Kidney	7.6
HUVEC starved	0.0		

General_screening_panel_v1.4 Summary: Ag2957 Expression of the NOV3b gene is restricted to placenta, fetal kidney and liver. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel. In addition, this gene shows no or very low expression in the cancer cell lines used in this panel. Thus, the absence of expression could potentially be used as a diagnostic marker for cancer.

Panel 1.3D Summary: Ag2957 Expression of the NOV3b gene is restricted to kidney, spinal cord and liver. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel. This gene encodes a putative claudin. Claudins are components of tight junction strands. Thus, this specific pattern of expression may indicate that this gene product is involved in the formation of TJ strands in these tissues.

Among the CNS regions on this panel, this tight junction protein is expressed only in the spinal cord and may be involved in the blood brain barrier in this region. This molecule may therefore be of utility in the treatment of spinal cord injury. Growth factors such as BDNF and NGF have been shown in animal models to enhance repair after spinal crush injury; however in the clinical condition it is hard to administer protein therapeutics due to the blood brain barrier. The selective downregulation of this molecule may therefore increase the amount of protein crossing the blood brain barrier in the spinal cord, while not hampering its function in the rest of the CNS.

In addition, this gene shows no or very low expression in the cancer cell lines used in this panel. Thus, the absence of expression could potentially be used as a diagnostic marker for cancer.

Panel 2D Summary: Ag2957 The NOV3b gene is consistently expressed in the normal kidney samples (CTs=32-33) but not in the adjacent kidney tumors. This result is in agreement with the expression in the previous panels. Thus, absence of expression of this gene could be used as a diagnostic marker for kidney cancer. Furthermore, therapeutic modulation of the function or expression of this gene may be a possible treatment for this cancer.

Panel 4D Summary: Ag2957 The expression of the NOV3b transcript is restricted to the thymus (CT=32.1) but not in T cells. Thus, expression of this transcript could be used as a marker for this tissue.

NOV1a, NOV1d, NOV1c, and NOV1b

THE UNIVERSITY OF CHICAGO PRESS

5

Table DB. Probe Name Ag2956

Table DC. General screening panel v1.4

909

Uterus Pool	0.0	0.0	Colon ca. SW1116	0.0	0.0
Ovarian ca. OVCAR-3	62.4	0.0	Colon ca. Colo-205	0.0	0.0
Ovarian ca. SK-OV-3	0.0	7.9	Colon ca. SW-48	0.0	0.0
Ovarian ca. OVCAR-4	69.7	0.0	Colon Pool	99.3	0.0
Ovarian ca. OVCAR-5	58.2	0.0	Small Intestine Pool	0.0	10.1
Ovarian ca. IGROV-1	18.7	0.0	Stomach Pool	25.9	10.3
Ovarian ca. OVCAR-8	0.0	0.0	Bone Marrow Pool	28.3	0.0
Ovary	0.0	0.0	Fetal Heart	0.0	0.0
Breast ca. MCF-7	78.5	0.0	Heart Pool	0.0	0.0
Breast ca. MDA-MB-231	0.0	0.0	Lymph Node Pool	0.0	0.0
Breast ca. BT 549	0.0	0.0	Fetal Skeletal Muscle	22.5	1.7
Breast ca. T47D	25.5	0.0	Skeletal Muscle Pool	62.0	26.2
Breast ca. MDA-N	57.8	0.0	Spleen Pool	59.5	0.0
Breast Pool	0.0	0.0	Thymus Pool	23.8	11.1
Trachea	0.0	0.0	CNS cancer (glio/astro) U87-MG	0.0	0.0
Lung	27.9	0.0	CNS cancer (glio/astro) U-118-MG	0.0	9.7
Fetal Lung	0.0	0.0	CNS cancer (neuro;met) SK-N-AS	17.0	0.0
Lung ca. NCI-N417	0.0	0.0	CNS cancer (astro) SF-539	0.0	0.0
Lung ca. LX-1	0.0	0.0	CNS cancer (astro) SNB-75	0.0	0.0
Lung ca. NCI-H146	0.0	0.0	CNS cancer (glio) SNB-19	0.0	0.0
Lung ca. SHP-77	0.0	0.0	CNS cancer (glio) SF-295	0.0	0.0
Lung ca. A549	28.7	0.0	Brain (Amygdala) Pool	0.0	0.0

Lung ca. NCI-H526	0.0	6.0	Brain (cerebellum)	56.3	3.3
Lung ca. NCI-H23	28.1	11.0	Brain (fetal)	27.5	100.0
Lung ca. NCI-H460	100.0	7.4	Brain (Hippocampus) Pool	0.0	0.0
Lung ca. HOP-62	0.0	0.0	Cerebral Cortex Pool	0.0	0.0
Lung ca. NCI-H522	0.0	0.0	Brain (Substantia nigra) Pool	26.8	0.0
Liver	0.0	0.0	Brain (Thalamus) Pool	0.0	0.0
Fetal Liver	54.7	0.0	Brain (whole)	58.2	6.2
Liver ca. HepG2	0.0	0.0	Spinal Cord Pool	0.0	0.0
Kidney Pool	0.0	8.0	Adrenal Gland	0.0	0.0
Fetal Kidney	0.0	0.0	Pituitary gland Pool	25.3	5.4
Renal ca. 786-0	0.0	0.0	Salivary Gland	0.0	0.0
Renal ca. A498	0.0	0.0	Thyroid (female)	0.0	0.0
Renal ca. ACHN	0.0	0.0	Pancreatic ca. CAPAN2	85.3	0.0
Renal ca. UO-31	0.0	0.0	Pancreas Pool	0.0	0.0

Table DD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2954, Run 164329620	Rel. Exp.(%) Ag2956, Run 164401746	Tissue Name	Rel. Exp.(%) Ag2954, Run 164329620	Rel. Exp.(%) Ag2956, Run 164401746
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	48.3	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	20.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC	0.0	0.0

			none		
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microvascular Dermal EC TNFalpha + IL- 1beta	13.2	0.0
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	38.7	0.0
Primary Tr1 rest	17.0	0.0	Small airway epithelium TNFalpha + IL- 1beta	0.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	19.1
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL- 1beta	0.0	0.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL- 1beta	0.0	0.0
LAK cells IL-2	0.0	0.0	Liver cirrhosis	72.2	100.0
LAK cells IL-2+IL- 12	0.0	0.0	Lupus kidney	0.0	0.0
LAK cells IL-	11.3	0.0	NCI-H292 none	0.0	0.0

2+IFN gamma					
LAK cells IL-2+ IL-18	0.0	0.0	NCI-H292 IL-4	14.5	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-9	0.0	0.0
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	0.0	0.0
Two Way MLR 3 day	0.0	0.0	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 5 day	0.0	0.0	HPAEC none	0.0	0.0
Two Way MLR 7 day	24.7	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	6.9	0.0	Lung fibroblast none	100.0	0.0
PBMC PWM	15.3	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	9.7	32.8
Ramos (B cell) ionomycin	10.7	0.0	Lung fibroblast IL-13	0.0	0.0
B lymphocytes PWM	13.0	0.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes CD40L and IL-4	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	13.1	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	34.2	26.8
Dendritic cells anti-CD40	0.0	0.0	IBD Colitis 2	34.9	0.0
Monocytes rest	0.0	0.0	IBD Crohn's	20.0	0.0
Monocytes LPS	0.0	0.0	Colon	0.0	0.0
Macrophages rest	13.0	0.0	Lung	0.0	0.0
Macrophages LPS	24.7	0.0	Thymus	12.2	0.0
HUVEC none	0.0	0.0	Kidney	0.0	0.0
HUVEC starved	40.1	0.0			

CNS_neurodegeneration_v1.0 Summary: Ag2954/Ag2956 Expression of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

General_screening_panel_v1.4 Summary: Ag2954 The NOV1a gene is expressed at a very low level or not at all in most of the cancer cell lines on this panel. Very low expression in cell lines from pancreatic, lung, breast and ovarian cancers suggests that it may be involved in these cancers.

Ag2956 This gene is a member of the claudin family of proteins, and is only expressed in the fetal brain. It may be involved in the process of axonal growth or targeting and synaptogenesis (specifically in the development of tight junctions between neurons and other cell types). Therefore, this gene product may be of therapeutic benefit in the treatment of neuronal loss in clinical conditions such as head trauma or stroke where increased compensatory synaptogenesis is desirable.

Panel 1.3D Summary: Ag2954/Ag2956 Expression of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2954/Ag 2956 Two experiments with two different sets of primers show low but significant levels of expression of this transcript in liver cirrhosis, dermal and lung fibroblasts and endothelium. Thus, the NOV11 transcript may serve as a marker for these tissues and play a role in maintaining the integrity of these tissues.

NOV2

Expression of gene NOV2 was assessed using the primer-probe set Ag2958, described in Table EA. Results of the RTQ-PCR runs are shown in Table EB.

Table EA. Probe Name Ag2958

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gttgtcagggtagagcagaaga-3'	22	341	1015
Probe	TET-5'-ccaccaagaaaccttttgcaataaaa-3'-TAMRA	26	363	1016
Reverse	5'-taccttcctctctctggtttc-3'	22	395	1017

Table EB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2958, Run 167809085	Tissue Name	Rel. Exp.(%) Ag2958, Run 167809085
Liver adenocarcinoma	0.0	Kidney (fetal)	0.4
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0

Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.6	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.4
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	2.0
Thymus	0.0	Ovarian ca. OVCAR- 4	0.0
Spleen	0.0	Ovarian ca. OVCAR-	0.0

		5	
Lymph node	0.0	Ovarian ca. OVCA-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	100.0	Adipose	0.0

Panel 1.3D Summary: Ag2958 Expression of the NOV2 is restricted to the kidney (CT=31.8). In addition, this gene is expressed at higher levels in adult kidney when compared to expression in fetal kidney (CT value = 40). Thus, this gene product may be useful for the differentiation of adult and fetal kidney tissue. This highly specific expression pattern also suggests that this gene product may be a small molecule drug for the treatment of diseases of the kidney.

Panel 4D Summary: Ag2958 Expression of the NOV2 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

10 NOV10

Expression of gene NOV10, also known as CG55964-02, was assessed using the primer-probe set Ag2857, described in Table FA.

Table FA. Probe Name Ag2857

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-catttgccagagatttcttttg-3'	22	300	1018
Probe	TET-5'-caaattgtggcttattcattcattccagg-3'-TAMRA	28	336	1019
Reverse	5'-agaaggatacccgattcaattg-3'	22	364	1020

CNS_neurodegeneration_v1.0 Summary: Ag2857 Expression of the NOV10 gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

5 **Panel 1.3D Summary:** Ag2857 Expression of the NOV10 gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

Panel 2.2 Summary: Ag2857 Expression of the NOV10 gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

10 **Panel 4D Summary:** Ag2857 Expression of the NOV10 gene is low/undetectable in all samples on this panel. (CTs>35). (Data not shown.)

NOV11

Expression of gene NOV11 was assessed using the primer-probe set Ag2858, described in Table GA.

Table GA. Probe Name Ag2858

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cttaagtcagtcctggcagttg-3'	22	674	1021
Probe	TET-5'-aaattatttcagacctgcattctccca-3'-TAMRA	26	716	1022
Reverse	5'-agaacacaaggacagcacagat-3'	22	743	1023

15

CNS_neurodegeneration_v1.0 Summary: Ag2858 Expressin of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

20 **Panel 1.3D Summary:** Ag2858 Expression of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 2.2 Summary: Ag2858 Expression of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 4D Summary: Ag2858 Expression of the NOV11 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

NOV12

5 Expression of gene NOV12 was assessed using the primer-probe set Ag2867,
described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB and HC.

Table HA. Probe Name Ag2867

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtcacaccttgtgactgcaa-3'	20	176	1024
Probe	TET-5'-acttacttgtgtgcagtgccatgcct-3'-TAMRA	26	196	1025
Reverse	5'-cattggaaacctttcaggaaa-3'	21	236	1026

Table HB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag2867, Run 208699903	Tissue Name	Rel. Exp.(%) Ag2867, Run 208699903
AD 1 Hippo	7.9	Control (Path) 3 Temporal Ctx	1.5
AD 2 Hippo	20.3	Control (Path) 4 Temporal Ctx	22.1
AD 3 Hippo	5.0	AD 1 Occipital Ctx	13.4
AD 4 Hippo	6.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	65.5	AD 3 Occipital Ctx	3.6
AD 6 Hippo	44.4	AD 4 Occipital Ctx	12.9
Control 2 Hippo	14.3	AD 5 Occipital Ctx	15.4
Control 4 Hippo	9.1	AD 6 Occipital Ctx	8.7
Control (Path) 3 Hippo	5.6	Control 1 Occipital Ctx	0.7
AD 1 Temporal Ctx	100.0	Control 2 Occipital Ctx	1.4
AD 2 Temporal Ctx	12.6	Control 3 Occipital Ctx	14.5
AD 3 Temporal Ctx	7.6	Control 4 Occipital Ctx	3.4
AD 4 Temporal Ctx	14.0	Control (Path) 1 Occipital Ctx	62.0
AD 5 Inf Temporal Ctx	55.9	Control (Path) 2 Occipital Ctx	8.1
AD 5 SupTemporal	71.2	Control (Path) 3	0.0

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	45.7	Control (Path) 4 Occipital Ctx	21.0
AD 6 Sup Temporal Ctx	16.5	Control 1 Parietal Ctx	0.9
Control 1 Temporal Ctx	0.9	Control 2 Parietal Ctx	8.3
Control 2 Temporal Ctx	5.4	Control 3 Parietal Ctx	4.4
Control 3 Temporal Ctx	2.5	Control (Path) 1 Parietal Ctx	40.1
Control 4 Temporal Ctx	6.3	Control (Path) 2 Parietal Ctx	20.7
Control (Path) 1 Temporal Ctx	39.8	Control (Path) 3 Parietal Ctx	5.2
Control (Path) 2 Temporal Ctx	24.8	Control (Path) 4 Parietal Ctx	21.3

Table HC. Panel 4D

	Tissue Name	Rel. Exp.(%) Ag2867, Run 164311002	Tissue Name	Rel. Exp.(%) Ag2867, Run 164311002
act	Secondary Th1	5.6	HUVEC IL-1beta	4.2
act	Secondary Th2	15.7	HUVEC IFN gamma	1.9
act	Secondary Tr1	11.8	HUVEC TNF alpha + IFN gamma	2.7
rest	Secondary Th1	3.9	HUVEC TNF alpha + IL4	2.9
rest	Secondary Th2	17.3	HUVEC IL-11	2.3
rest	Secondary Tr1	12.2	Lung Microvascular EC none	6.1
	Primary Th1 act	38.4	Lung Microvascular EC TNFalpha + IL-1beta	7.2
	Primary Th2 act	22.2	Microvascular Dermal EC none	6.7
	Primary Tr1 act	29.7	Microvascular Dermal EC TNFalpha + IL-1beta	3.9
	Primary Th1 rest	93.3	Bronchial epithelium TNFalpha + IL1beta	9.0
	Primary Th2 rest	55.9	Small airway epithelium none	1.7

Primary Tr1 rest	39.2	Small airway epithelium TNFalpha + IL-1beta	16.7
CD45RA CD4 lymphocyte act	6.0	Coronary artery SMC rest	3.2
CD45RO CD4 lymphocyte act	15.6	Coronary artery SMC TNFalpha + IL-1beta	1.3
CD8 lymphocyte act	16.0	Astrocytes rest	8.8
Secondary CD8 lymphocyte rest	17.2	Astrocytes TNFalpha + IL-1beta	6.2
Secondary CD8 lymphocyte act	13.6	KU-812 (Basophil) rest	0.7
CD4 lymphocyte none	22.4	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	24.0	CCD1106 (Keratinocytes) none	7.0
LAK cells rest	23.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.4
LAK cells IL-2	29.9	Liver cirrhosis	3.6
LAK cells IL-2+IL-12	32.8	Lupus kidney	1.1
LAK cells IL-2+IFN gamma	50.7	NCI-H292 none	17.8
LAK cells IL-2+IL-18	39.0	NCI-H292 IL-4	15.5
LAK cells PMA/ionomycin	4.1	NCI-H292 IL-9	13.9
NK Cells IL-2 rest	22.5	NCI-H292 IL-13	7.2
Two Way MLR 3 day	26.1	NCI-H292 IFN gamma	8.2
Two Way MLR 5 day	9.3	HPAEC none	2.7
Two Way MLR 7 day	12.0	HPAEC TNF alpha + IL-1 beta	2.7
PBMC rest	11.6	Lung fibroblast none	5.5
PBMC PWM	65.1	Lung fibroblast TNF alpha + IL-1 beta	2.8
PBMC PHA-L	17.1	Lung fibroblast IL-4	7.2
Ramos (B cell) none	13.9	Lung fibroblast IL-9	6.9
Ramos (B cell) ionomycin	23.8	Lung fibroblast IL-13	6.5

B lymphocytes PWM	73.2	Lung fibroblast IFN gamma	4.6
B lymphocytes CD40L and IL-4	100.0	Dermal fibroblast CCD1070 rest	12.7
EOL-1 dbcAMP	3.1	Dermal fibroblast CCD1070 TNF alpha	40.1
EOL-1 dbcAMP PMA/ionomycin	5.2	Dermal fibroblast CCD1070 IL-1 beta	6.0
Dendritic cells none	6.5	Dermal fibroblast IFN gamma	2.6
Dendritic cells LPS	3.1	Dermal fibroblast IL-4	5.8
Dendritic cells anti-CD40	4.2	IBD Colitis 2	2.7
Monocytes rest	8.5	IBD Crohn's	5.5
Monocytes LPS	5.0	Colon	22.5
Macrophages rest	4.2	Lung	6.7
Macrophages LPS	2.4	Thymus	6.7
HUVEC none	2.6	Kidney	92.7
HUVEC starved	12.7		

CNS_neurodegeneration_v1.0 Summary: Ag2867 The NOV12 gene represents a novel G-protein coupled receptor (GPCR) with expression in the brain. The GPCR family of receptors contains a large number of neurotransmitter receptors, including the dopamine, serotonin, α and β -adrenergic, acetylcholine muscarinic, histamine, peptide, and metabotropic glutamate receptors. GPCRs are excellent drug targets in various neurologic and psychiatric diseases. All antipsychotics have been shown to act at the dopamine D2 receptor; similarly novel antipsychotics also act at the serotonergic receptor, and often the muscarinic and adrenergic receptors as well. While the majority of antidepressants can be classified as selective serotonin reuptake inhibitors, blockade of the 5-HT1A and α 2 adrenergic receptors increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; further more the purinergic receptors have also been implicated as drug targets in the treatment of cerebral ischemia. The β -adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the α -adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

In addition, this GPCR is found to be upregulated in the temporal cortex of Alzheimer's disease patients. Blockade of this receptor may be of use in the treatment of this disease and decrease neuronal death.

References:

- 5 El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. *Br J Pharmacol* 2001 Sep;134(1):68-77
1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg(-1), i.p.) and KW 6002 (0.1 - 10 mg kg(-1), p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg(-1)) and ZM 241385 (15 - 60 mg kg(-1)) were effective in mice previously screened for having high immobility time, while SCH 58261 at 10 mg kg(-1) reduced immobility of mice that were selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg(-1) reduced the immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg(-1) by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg(-1) i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg(-1) i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants. Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.
- Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. *Clin Psychiatry* 2001;62 Suppl 15:12-7

Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT_{1A} (cell body) and 5-HT_{1B} (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term treatment desensitizes the inhibitory 5-HT₁ autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT_{1A} receptors, to SSRI treatment decouples the feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha₁-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha₂-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha₂-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. *Expert Opin Investig Drugs* 1999 Nov;8(11):1837-1848

The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis. NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified

can be divided into five main categories depending on their chemical structure: indoles, tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats.

5 Neuroreport 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg
10 either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a
15 potentially useful biological target for the reduction of brain injury.

Panel 1.3D Summary: Ag2867 Results from one experiment with the NOV12 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 2.2 Summary: Ag2867 Results from one experiment with the NOV12 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

20 **Panel 4D Summary:** Ag2867 Expression of the NOV12 gene is widespread among samples in this panel, with highest expression in B lymphocytes stimulated with CD40L and IL-4 (CT=31.1).

This transcript is also highly expressed in activated B cells and primary resting Th1 and Th2 T cells. The expression of this transcript in PBMC treated with the B cell mitogen, PWM, confirms the importance of CG54575-01 gene expression in activated B cells. In
25 addition, this transcript is also abundantly expressed on primary resting Th1 cells (to a lesser degree on primary resting Th2 cells). Therefore, it appears that this gene, encoding a GPCR homolog, is a potential new member of the chemokine receptor family. The expression of this protein in activated B cells suggests a role for this protein in their trafficking to appropriate
30 sites where they can fully activate antigen specific T cells. Thus, the protein encoded by this gene is likely to participate in the development of immune or inflammatory reactions.

In addition, the high expression of this gene in the kidney suggests that the putative GPCR encoded for by this gene could allow cells within the kidney to respond to specific microenvironmental signals (For example, ref. 1). Therefore, antibody or small molecule

therapies designed with the protein encoded for by this gene could modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

References:

5 References:

Mark M.D., Wittemann S., Herlitz S. (2000) G protein modulation of recombinant P/Q-type calcium channels by regulators of G protein signalling proteins. J. Physiol. 528 Pt 1: 65-77.

1. Fast synaptic transmission is triggered by the activation of presynaptic Ca²⁺ channels which can be inhibited by Gbetagamma subunits via G protein-coupled receptors (GPCR). Regulators of G protein signalling (RGS) proteins are GTPase-accelerating proteins (GAPs), which are responsible for >100-fold increases in the GTPase activity of G proteins and might be involved in the regulation of presynaptic Ca²⁺ channels. In this study we investigated the effects of RGS2 on G protein modulation of recombinant P/Q-type channels expressed in a human embryonic kidney (HEK293) cell line using whole-cell recordings. 2. RGS2 markedly accelerates transmitter-mediated inhibition and recovery from inhibition of Ba²⁺ currents (IBa) through P/Q-type channels heterologously expressed with the muscarinic acetylcholine receptor M2 (mAChR M2). 3. Both RGS2 and RGS4 modulate the prepulse facilitation properties of P/Q-type Ca²⁺ channels. G protein reinhibition is accelerated, while release from inhibition is slowed. These kinetics depend on the availability of G protein alpha and betagamma subunits which is altered by RGS proteins. 4. RGS proteins unmask the Ca²⁺ channel beta subunit modulation of Ca²⁺ channel G protein inhibition. In the presence of RGS2, P/Q-type channels containing the beta2a and beta3 subunits reveal significantly altered kinetics of G protein modulation and increased facilitation compared to Ca²⁺ channels coexpressed with the beta1b or beta4 subunit.

PMID: 11018106

NOV13a and NOV13b

Expression of gene NOV13a and variant NOV13b was assessed using the primer-probe set Ag2869, described in Table IA. Results of the RTQ-PCR runs are shown in Table IB.

30

Table IA. Probe Name Ag2869

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgtctgtggttagacaccacctt-3'	22	475	1027
Probe	TET-5'-ctgaggctaccctaccgaggcagtaa-3'-	26	501	1028

	TAMRA			
Reverse	5' - cacaaaagaaatgagcaatgct - 3'	22	528	1029

Table IB. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2869, Run 164311007	Tissue Name	Rel. Exp.(%) Ag2869, Run 164311007
Secondary Th1 act	5.5	HUVEC IL-1beta	1.7
Secondary Th2 act	68.3	HUVEC IFN gamma	9.5
Secondary Tr1 act	33.9	HUVEC TNF alpha + IFN gamma	5.6
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	8.9
Secondary Th2 rest	7.3	HUVEC IL-11	5.5
Secondary Tr1 rest	3.0	Lung Microvascular EC none	11.0
Primary Th1 act	40.9	Lung Microvascular EC TNFalpha + IL-1beta	12.0
Primary Th2 act	19.1	Microvascular Dermal EC none	3.1
Primary Tr1 act	51.4	Microvascular Dermal EC TNFalpha + IL-1beta	1.2
Primary Th1 rest	23.5	Bronchial epithelium TNFalpha + IL1beta	2.4
Primary Th2 rest	11.8	Small airway epithelium none	0.8
Primary Tr1 rest	17.2	Small airway epithelium TNFalpha + IL-1beta	9.6
CD45RA CD4 lymphocyte act	1.6	Coronary artery SMC rest	5.5
CD45RO CD4 lymphocyte act	12.8	Coronary artery SMC TNFalpha + IL-1beta	2.5
CD8 lymphocyte act	1.0	Astrocytes rest	6.3
Secondary CD8 lymphocyte rest	2.7	Astrocytes TNFalpha + IL-1beta	1.8
Secondary CD8 lymphocyte act	3.4	KU-812 (Basophil) rest	1.8
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	21.6
2ry Th1/Th2/Tr1 _anti- CD95 CH11	1.5	CCD1106 (Keratinocytes) none	6.3
LAK cells rest	0.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.0
LAK cells IL-2	0.0	Liver cirrhosis	5.4
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.9

LAK cells IL-2+IFN gamma	3.1	NCI-H292 none	34.2
LAK cells IL-2+ IL-18	2.4	NCI-H292 IL-4	46.3
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	29.7
NK Cells IL-2 rest	1.3	NCI-H292 IL-13	22.8
Two Way MLR 3 day	0.6	NCI-H292 IFN gamma	25.2
Two Way MLR 5 day	1.1	HPAEC none	18.0
Two Way MLR 7 day	3.1	HPAEC TNF alpha + IL-1 beta	19.9
PBMC rest	0.0	Lung fibroblast none	0.8
PBMC PWM	4.1	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	9.9	Lung fibroblast IL-4	8.9
Ramos (B cell) none	50.7	Lung fibroblast IL-9	7.9
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	4.9
B lymphocytes PWM	1.1	Lung fibroblast IFN gamma	4.9
B lymphocytes CD40L and IL-4	3.4	Dermal fibroblast CCD1070 rest	42.6
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	39.8
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	13.4
Dendritic cells none	2.1	Dermal fibroblast IFN gamma	2.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	2.6
Dendritic cells anti-CD40	0.6	IBD Colitis 2	1.1
Monocytes rest	0.0	IBD Crohn's	2.2
Monocytes LPS	0.0	Colon	3.2
Macrophages rest	4.8	Lung	1.0
Macrophages LPS	2.4	Thymus	4.4
HUVEC none	16.5	Kidney	16.3
HUVEC starved	17.6		

Panel 1.3D Summary: Ag2869 Expression of the NOV13a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

5

Panel 2.2 Summary: Ag2869 Expression of the NOV13a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2869 Expression of the NOV13a gene is widespread among the samples in this panel, with highest expression in the B cell line Ramos treated with ionomycin (CT=31.1). Lower but still significant levels of expression are seen in untreated Ramos B cells. B cells represent a principle component of immunity and contribute to the immune response in a number of important functional roles, including antibody production. For example, production of antibodies against self-antigens is a major component in autoimmune disorders such a systemic lupus erythematosus, with B cells playing a major role. Since B cells play an important role in autoimmunity, inflammatory processes and inflammatory cascades, therapeutic modulation of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, allergies, chronic obstructive pulmonary disease, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, psoriasis, osteoarthritis, and other autoimmune disorders including systemic lupus erythematosus.

Significant levels of expression are also seen in IL-4, IL-9, IL-13 and IFN gamma activated-NCI-H292 mucoepidermoid cells as well as untreated NCI-H292 cells. Moderate expression is also detected in both treated and untreated human pulmonary aortic endothelial cells. The expression of this gene in cells derived from or within the lung suggests that this gene may be involved in normal conditions as well as pathological and inflammatory lung disorders that include chronic obstructive pulmonary disease, asthma, allergy and emphysema.

NOV14

20 Expression of gene NOV14 was assessed using the primer-probe set Ag2870,
described in Table JA. Results of the RTQ-PCR runs are shown in Tables JB and JC.

Table JA. Probe Name Ag2870

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ttctcctgagagcaagaagtg-3'	22	823	1030
Probe	TET-5'-tgtgactcccatgttgaaacccatta-3'-TAMRA	26	865	1031
Reverse	5'-tcttcacctcgctattttctcaa-3'	22	899	1032

Table JB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2870, Run 167646334	Tissue Name	Rel. Exp.(%) Ag2870, Run 167646334
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	2.5
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0

Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	7.5
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	6.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	20.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	3.8
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.0
Thymus	8.7	Ovarian ca. OVCAR- 4	0.0
Spleen	0.0	Ovarian ca. OVCAR-	28.9

		5	
Lymph node	0.0	Ovarian ca. OVCAR-8	13.4
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	21.3
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	100.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	5.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	33.9
Kidney	0.0	Adipose	0.0

Table JC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2870, Run 164328103	Tissue Name	Rel. Exp.(%) Ag2870, Run 164328103
Secondary Th1 act	9.7	HUVEC IL-1beta	9.3
Secondary Th2 act	30.4	HUVEC IFN gamma	11.6
Secondary Tr1 act	25.2	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.2	HUVEC TNF alpha + IL4	8.2
Secondary Th2 rest	1.9	HUVEC IL-11	0.0
Secondary Tr1 rest	2.0	Lung Microvascular EC none	11.5
Primary Th1 act	16.7	Lung Microvascular EC TNFalpha + IL-1beta	3.6
Primary Th2 act	12.3	Microvascular Dermal EC none	0.0
Primary Tr1 act	29.1	Microsvascular Dermal EC	2.6

		TNFalpha + IL-1beta	
Primary Th1 rest	9.3	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	1.1	Small airway epithelium none	0.0
Primary Tr1 rest	4.5	Small airway epithelium TNFalpha + IL-1beta	13.3
CD45RA CD4 lymphocyte act	3.2	Coronary artery SMC rest	1.7
CD45RO CD4 lymphocyte act	6.5	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	1.6	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	1.4	KU-812 (Basophil) rest	2.4
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	8.7
2ry Th1/Th2/Tr1_anti- CD95 CH11	4.7	CCD1106 (Keratinocytes) none	6.6
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	1.2
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.6	NCI-H292 none	30.8
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	32.5
LAK cells PMA/ionomycin	0.4	NCI-H292 IL-9	26.8
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	11.3
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	21.3
Two Way MLR 5 day	0.0	HPAEC none	9.2
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	8.0
PBMC rest	0.0	Lung fibroblast none	1.9
PBMC PWM	0.9	Lung fibroblast TNF alpha + IL-1 beta	1.8
PBMC PHA-L	5.0	Lung fibroblast IL-4	7.5
Ramos (B cell) none	36.3	Lung fibroblast IL-9	7.0
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	0.5
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	1.4
B lymphocytes CD40L and IL-4	3.1	Dermal fibroblast CCD1070 rest	26.8

EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	23.2
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	14.6
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	2.3
Dendritic cells anti- CD40	0.0	IBD Colitis 2	0.4
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	1.8
Macrophages rest	3.5	Lung	0.0
Macrophages LPS	0.0	Thymus	0.3
HUVEC none	12.9	Kidney	8.1
HUVEC starved	22.7		

Panel 1.3D Summary: Ag2870 Expression of the NOV14 gene is restricted to a sample derived from a colon cancer cell line (CT=34.4). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of colon cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon cancer.

Panel 2.2 Summary: Ag2870 Expression of the NOV14 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2870 Expression of the NOV14 gene is highest in the B cell line Ramos treated with ionomycin (CT=30.2). Lower but still significant levels of expression are seen in untreated Ramos B cells. B cells represent a principle component of immunity and contribute to the immune response in a number of important functional roles, including antibody production. For example, production of antibodies against self-antigens is a major component in autoimmune disorders such a systemic lupus erythematosus, with B cells playing a major role. Since B cells play an important role in autoimmunity, inflammatory processes and inflammatory cascades, therapeutic modulation of this gene product may reduce or eliminate the symptoms of patients suffering from asthma, allergies, chronic obstructive pulmonary disease, emphysema, Crohn's disease, ulcerative colitis, rheumatoid arthritis, psoriasis, osteoarthritis, and other autoimmune disorders including systemic lupus erythematosus.

Significant levels of expression are also seen in IL-4, IL-9, IL-13, IFN gamma activated and untreated NCI-H292 mucoepidermoid cells, IL-4, IL-9, IL-13 and IFN gamma activated lung fibroblasts, human pulmonary aortic endothelial cells (treated and untreated),

treated small airway epithelium and lung microvascular endothelial cells (treated and untreated). The expression of this gene in cells derived from or within the lung further suggests that this gene may be involved in normal conditions as well as pathological and inflammatory lung disorders that include chronic obstructive pulmonary disease, asthma, allergy and emphysema.

NOV15a and NOV15b

Expression of gene NOV15a and variant NOV15b was assessed using the primer-probe set Ag2875, described in Table KA. Results of the RTQ-PCR runs are shown in Table KB.

Table KA. Probe Name Ag2875

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaacatcatctcctaccctgaa-3'	22	268	1033
Probe	TET-5'-tgcatgactcagctctacttcttctctcg-3'-TAMRA	28	290	1034
Reverse	5'-atgtgacactctgcaatagcaa-3'	22	321	1035

10

Table KB. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2875, Run 164311029	Tissue Name	Rel. Exp.(%) Ag2875, Run 164311029
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	3.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0

CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	24.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	5.1	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	72.2
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	20.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	1.4
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	3.2
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	3.1	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	33.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	2.4	Dermal fibroblast IL-4	1.9

Dendritic cells anti-CD40	100.0	IBD Colitis 2	16.3
Monocytes rest	0.0	IBD Crohn's	6.7
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	44.1	Lung	21.8
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	3.1	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2875 Results from one experiment with the NOV15a gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 2.2 Summary: Ag2875 Expression of the NOV15a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2875 Highest expression of the NOV15a is in anti-CD40 treated dendritic cells (CT=33.2), with much lower expression in untreated dendritic cells. Thus, this gene product may be important in dendritic cell activation. Significant expression of this gene is also seen in liver cirrhosis. This gene encodes a putative GPCR; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative GPCR could also be used for the diagnosis of liver cirrhosis. In addition, significant expression of this gene is seen in resting macrophages. The putative GPCR encoded for by this transcript may therefore be important in macrophage detection of chemokine gradients and trafficking into specific sites within a tissue and in activation. Antibody or protein therapeutics designed against the protein encoded for by this transcript could reduce or inhibit inflammation in asthma, emphysema, allergy, psoriasis, arthritis, or any other condition in which macrophage localization/activation is important.

NOV16A: Olfactory Receptor

Expression of gene NOV16a was assessed using the primer-probe set Ag2876, described in Table LA. Results of the RTQ-PCR runs are shown in Tables LB and LC.

Table LA. Probe Name Ag2876

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acatcatctcctaccctgaatg-3'	22	273	1036
Probe	TET-5'-catgactcagctttacttcttctcatt-3'-TAMRA	28	295	1037
Reverse	5'-tacagccaacatgtgacactct-3'	22	334	1038

Table LB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2876, Run 167646343	Tissue Name	Rel. Exp.(%) Ag2876, Run 167646343
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	1.7	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	1.4	Breast ca.* (pl.ef) MCF-7	19.2
glioma SF-295	2.6	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	100.0
Heart	0.0	Breast ca. BT-549	0.0

Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	2.1	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	2.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	3.6
Colon ca. HCT-116	0.0	Testis	8.5
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	24.7
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table LC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2876, Run 164328133	Tissue Name	Rel. Exp.(%) Ag2876, Run 164328133
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0

Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	2.1
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	9.7	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	30.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	8.8
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0

Ramos (B cell) none	0.0	Lung fibroblast IL-9	4.8
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	15.6
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	5.4
Dendritic cells none	10.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	19.3
Dendritic cells anti-CD40	18.8	IBD Colitis 2	15.3
Monocytes rest	0.0	IBD Crohn's	2.2
Monocytes LPS	20.6	Colon	4.1
Macrophages rest	14.2	Lung	21.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2876 Expression of the NOV16a gene is restricted to a sample derived from a breast cancer cell line (CT=32.5). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of breast cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of breast cancer.

Panel 4D Summary: Ag2876 Significant expression of the NOV16a gene is detected in a liver cirrhosis sample (CT = 33.5). Furthermore, expression of this gene is not detected in normal liver in Panel 1.3D, suggesting that its expression is unique to liver cirrhosis. This gene encodes a putative GPCR; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative GPCR could also be used for the diagnosis of liver cirrhosis.

NOV17a and NOV17b

Expression of gene NOV17a and variant NOV17b was assessed using the primer-probe set Ag2969, described in Table MA. Results of the RTQ-PCR runs are shown in Tables MB and MC.

Table MA. Probe Name Ag2969

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cacccatgtatttctctgcttag-3'	22	184	1039
Probe	TET-5'-tcagctctccctcattgacctaaatt-3'- TAMRA	26	206	1040
Reverse	5'-tcagaagccatcttaggaacaa-3'	22	244	1041

Table MB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2969, Run 209778982	Tissue Name	Rel. Exp.(%) Ag2969, Run 209778982
AD 1 Hippo	0.5	Control (Path) 3 Temporal Ctx	0.5
AD 2 Hippo	9.1	Control (Path) 4 Temporal Ctx	23.7
AD 3 Hippo	0.7	AD 1 Occipital Ctx	4.2
AD 4 Hippo	1.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	1.9
AD 6 Hippo	24.3	AD 4 Occipital Ctx	9.5
Control 2 Hippo	4.7	AD 5 Occipital Ctx	17.8
Control 4 Hippo	2.1	AD 6 Occipital Ctx	14.5
Control (Path) 3 Hippo	1.1	Control 1 Occipital Ctx	0.8
AD 1 Temporal Ctx	2.0	Control 2 Occipital Ctx	29.7
AD 2 Temporal Ctx	15.5	Control 3 Occipital Ctx	13.9
AD 3 Temporal Ctx	2.4	Control 4 Occipital Ctx	1.0
AD 4 Temporal Ctx	9.7	Control (Path) 1 Occipital Ctx	65.5
AD 5 Inf Temporal Ctx	91.4	Control (Path) 2 Occipital Ctx	4.8
AD 5 Sup Temporal Ctx	27.9	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	42.9	Control (Path) 4 Occipital Ctx	13.6
AD 6 Sup Temporal Ctx	39.2	Control 1 Parietal Ctx	1.2
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	28.9
Control 2 Temporal Ctx	11.8	Control 3 Parietal Ctx	9.6
Control 3 Temporal	9.3	Control (Path) 1	48.0

Ctx		Parietal Ctx	
Control 3 Temporal Ctx	2.5	Control (Path) 2 Parietal Ctx	12.0
Control (Path) 1 Temporal Ctx	38.4	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	12.2	Control (Path) 4 Parietal Ctx	41.2

Table MC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2969, Run 164402668	Tissue Name	Rel. Exp.(%) Ag2969, Run 164402668
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	3.0	HUVEC IFN gamma	2.8
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	6.1	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	4.1	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	10.5	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	13.6	Small airway epithelium none	0.0
Primary Tr1 rest	7.9	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	2.0
CD45RO CD4 lymphocyte act	7.5	Coronary artery SMC TNFalpha + IL-1beta	3.4
CD8 lymphocyte act	4.9	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	4.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	8.5
CD4 lymphocyte none	16.7	KU-812 (Basophil) PMA/ionomycin	51.1
2ry Th1/Th2/Tr1_anti-	3.2	CCD1106 (Keratinocytes)	0.0

CD95 CH11		none	
LAK cells rest	23.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	29.3	Liver cirrhosis	18.6
LAK cells IL-2+IL-12	22.2	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	28.5	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	21.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	7.2	NCI-H292 IL-13	0.0
Two Way MLR 3 day	51.4	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	18.4	HPAEC none	0.0
Two Way MLR 7 day	2.9	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	45.7	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	13.8	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	5.4	Lung fibroblast IL-13	0.0
B lymphocytes PWM	12.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	39.2	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	1.1	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	4.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	5.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti- CD40	6.4	IBD Colitis 2	11.1
Monocytes rest	1.5	IBD Crohn's	0.0
Monocytes LPS	10.7	Colon	0.0
Macrophages rest	11.4	Lung	2.4
Macrophages LPS	3.7	Thymus	0.0
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag2969 The NOV17a gene represents a novel G-protein coupled receptor (GPCR) with expression in the brain. The GPCR family of

receptors contains a large number of neurotransmitter receptors, including the dopamine, serotonin, α and β -adrenergic, acetylcholine muscarinic, histamine, peptide, and metabotropic glutamate receptors. GPCRs are excellent drug targets in various neurologic and psychiatric diseases. All antipsychotics have been shown to act at the dopamine D2 receptor; similarly novel antipsychotics also act at the serotonergic receptor, and often the muscarinic and adrenergic receptors as well. While the majority of antidepressants can be classified as selective serotonin reuptake inhibitors, blockade of the 5-HT1A and α_2 adrenergic receptors increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; furthermore the purinergic receptors have also been implicated as drug targets in the treatment of cerebral ischemia. The β -adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the α -adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

In addition, this panel shows that this GPCR is upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, blockade of this receptor may be of use in the treatment of this disease and decrease neuronal death.

References:

El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. *Br J Pharmacol* 2001 Sep;134(1):68-77

1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg⁻¹, i.p.) and KW 6002 (0.1 - 10 mg kg⁻¹, p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg⁻¹) and ZM 241385 (15 - 60 mg kg⁻¹) were effective in mice previously screened for having

high immobility time, while SCH 58261 at 10 mg kg⁻¹) reduced immobility of mice that were selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg⁻¹) reduced the immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg⁻¹) by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg⁻¹) i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg⁻¹) i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants. Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.

Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. Clin Psychiatry 2001;62 Suppl 15:12-7

Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT_{1A} (cell body) and 5-HT_{1B} (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term treatment desensitizes the inhibitory 5-HT₁ autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT_{1A} receptors, to SSRI treatment decouples the feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha1-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha2-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha2-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. Expert Opin Investig Drugs 1999 Nov;8(11):1837-1848

The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis. NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified can be divided into five main categories depending on their chemical structure: indoles, tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats. *Neuroreport* 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a potentially useful biological target for the reduction of brain injury.

Panel 1.3D Summary: Ag2878 Expression of the NOV17a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2878 Expression of the NOV17a gene is restricted to a few samples in this panel, with highest expression in the kidney (CT=33.1). Thus, the putative GPCR encoded for by this gene could allow cells within the kidney to respond to specific microenvironmental signals (For example, ref. 1). Therefore, antibody or small molecule therapies designed with the protein encoded for by this gene could modulate kidney function

and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

Furthermore, significant levels of expression are also seen in the PMA and ionomycin treated basophil cell line KU-812. GPCR-type receptors are important in multiple physiological responses mediated by basophils (ref. 2). Therefore, antibody or small molecule therapies designed with the protein encoded for by this gene could also block or inhibit inflammation or tissue damage due to basophil activation in response to asthma, allergies, hypersensitivity reactions, psoriasis, and viral infections.

References:

1. Mark M.D., Wittemann S., Herlitz S. (2000) G protein modulation of recombinant P/Q-type calcium channels by regulators of G protein signalling proteins. *J. Physiol.* 528 Pt 1: 65-77.

1. Fast synaptic transmission is triggered by the activation of presynaptic Ca²⁺ channels which can be inhibited by Gbetagamma subunits via G protein-coupled receptors (GPCR). Regulators of G protein signalling (RGS) proteins are GTPase-accelerating proteins (GAPs), which are responsible for >100-fold increases in the GTPase activity of G proteins and might be involved in the regulation of presynaptic Ca²⁺ channels. In this study we investigated the effects of RGS2 on G protein modulation of recombinant P/Q-type channels expressed in a human embryonic kidney (HEK293) cell line using whole-cell recordings. 2. RGS2 markedly accelerates transmitter-mediated inhibition and recovery from inhibition of Ba²⁺ currents (IBa) through P/Q-type channels heterologously expressed with the muscarinic acetylcholine receptor M2 (mAChR M2). 3. Both RGS2 and RGS4 modulate the prepulse facilitation properties of P/Q-type Ca²⁺ channels. G protein reinhibition is accelerated, while release from inhibition is slowed. These kinetics depend on the availability of G protein alpha and betagamma subunits which is altered by RGS proteins. 4. RGS proteins unmask the Ca²⁺ channel beta subunit modulation of Ca²⁺ channel G protein inhibition. In the presence of RGS2, P/Q-type channels containing the beta2a and beta3 subunits reveal significantly altered kinetics of G protein modulation and increased facilitation compared to Ca²⁺ channels coexpressed with the beta1b or beta4 subunit.

PMID: 11018106

2. Heinemann A., Hartnell A., Stubbs V.E., Murakami K., Soler D., LaRosa G., Askenase P.W., Williams T.J., Sabroe I. (2000) Basophil responses to chemokines are regulated by both sequential and cooperative receptor signaling. *J. Immunol.* 165: 7224-7233.

To investigate human basophil responses to chemokines, we have developed a sensitive assay that uses flow cytometry to measure leukocyte shape change as a marker of cell responsiveness. PBMC were isolated from the blood of volunteers. Basophils were identified as a single population of cells that stained positive for IL-3Ralpha (CDw123) and negative for HLA-DR, and their increase in forward scatter (as a result of cell shape change) in response to chemokines was measured. Shape change responses of basophils to chemokines were highly reproducible, with a rank order of potency: monocyte chemoattractant protein (MCP) 4 (peak at \approx eotaxin-2 = eotaxin-3 \approx eotaxin > MCP-1 = MCP-3 > macrophage-inflammatory protein-1alpha > RANTES = MCP-2 = IL-8. The CCR4-selective ligand macrophage-derived chemokine did not elicit a response at concentrations up to 10 nM. Blocking mAbs to CCR2 and CCR3 demonstrated that responses to higher concentrations (>10 nM) of MCP-1 were mediated by CCR3 rather than CCR2, whereas MCP-4 exhibited a biphasic response consistent with sequential activation of CCR3 at lower concentrations and CCR2 at 10 nM MCP-4 and above. In contrast, responses to MCP-3 were blocked only in the presence of both mAbs, but not after pretreatment with either anti-CCR2 or anti-CCR3 mAb alone. These patterns of receptor usage were different from those seen for eosinophils and monocytes. We suggest that cooperation between CCRs might be a mechanism for preferential recruitment of basophils, as occurs in tissue hypersensitivity responses in vivo.

PMID: 11120855

NOV17c

Expression of gene NOV17c, also known as CG56659-02, was assessed using the primer-probe set Ag2970, described in Table NA. Results of the RTQ-PCR runs are shown in Table NB.

Table NA. Probe Name Ag2970

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acatccatctccacacacctat-3'	22	90	1042
Probe	TET-5'-agtcagctctccctcattgacctaaa-3'-TAMRA	26	125	1043
Reverse	5'-taaaccatctttggaacaatgg-3'	22	162	1044

Table NB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2970, Run 211008706	Tissue Name	Rel. Exp.(%) Ag2970, Run 211008706
AD 1 Hippo	2.3	Control (Path) 3	2.1

		Temporal Ctx	
AD 2 Hippo	16.7	Control (Path) 4 Temporal Ctx	38.2
AD 3 Hippo	3.5	AD 1 Occipital Ctx	3.7
AD 4 Hippo	2.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	100.0	AD 3 Occipital Ctx	4.1
AD 6 Hippo	28.1	AD 4 Occipital Ctx	20.4
Control 2 Hippo	10.2	AD 5 Occipital Ctx	11.2
Control 4 Hippo	1.9	AD 6 Occipital Ctx	18.8
Control (Path) 3 Hippo	2.4	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	3.4	Control 2 Occipital Ctx	27.2
AD 2 Temporal Ctx	23.0	Control 3 Occipital Ctx	14.0
AD 3 Temporal Ctx	1.0	Control 4 Occipital Ctx	4.8
AD 4 Temporal Ctx	18.2	Control (Path) 1 Occipital Ctx	87.7
AD 5 Inf Temporal Ctx	99.3	Control (Path) 2 Occipital Ctx	9.9
AD 5 Sup Temporal Ctx	30.1	Control (Path) 3 Occipital Ctx	0.5
AD 6 Inf Temporal Ctx	47.0	Control (Path) 4 Occipital Ctx	13.5
AD 6 Sup Temporal Ctx	59.9	Control 1 Parietal Ctx	1.1
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	40.1
Control 2 Temporal Ctx	17.0	Control 3 Parietal Ctx	13.2
Control 3 Temporal Ctx	10.7	Control (Path) 1 Parietal Ctx	51.1
Control 4 Temporal Ctx	6.9	Control (Path) 2 Parietal Ctx	14.9
Control (Path) 1 Temporal Ctx	48.0	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	17.6	Control (Path) 4 Parietal Ctx	38.4

CNS_neurodegeneration_v1.0 Summary: Ag2970 The NOV17c gene represents a novel G-protein coupled receptor (GPCR) with expression in the brain. The GPCR family of receptors contains a large number of neurotransmitter receptors, including the dopamine,

serotonin, α and β -adrenergic, acetylcholine muscarinic, histamine, peptide, and metabotropic glutamate receptors. GPCRs are excellent drug targets in various neurologic and psychiatric diseases. All antipsychotics have been shown to act at the dopamine D2 receptor; similarly novel antipsychotics also act at the serotonergic receptor, and often the muscarinic and
5 adrenergic receptors as well. While the majority of antidepressants can be classified as selective serotonin reuptake inhibitors, blockade of the 5-HT1A and α_2 adrenergic receptors increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; further more the purinergic receptors have also been implicated as drug targets
10 in the treatment of cerebral ischemia. The β -adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the α -adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

In addition, this GPCR is upregulated in the temporal cortex of Alzheimer's disease
15 patients. Therefore, blockade of this receptor may be of use in the treatment of this disease and decrease neuronal death.

References:

El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on
20 pharmacology and A2A receptor knockout mice. *Br J Pharmacol* 2001 Sep;134(1):68-77

1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor
25 antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg⁻¹, i.p.) and KW 6002 (0.1 - 10 mg kg⁻¹,
30 p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg⁻¹) and ZM 241385 (15 - 60 mg kg⁻¹) were effective in mice previously screened for having high immobility time, while SCH 58261 at 10 mg kg⁻¹ reduced immobility of mice that were

selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg⁻¹) reduced the immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg⁻¹) by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg⁻¹) i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg⁻¹) i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants. Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.

Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. Clin Psychiatry 2001;62 Suppl 15:12-7

Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT_{1A} (cell body) and 5-HT_{1B} (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term treatment desensitizes the inhibitory 5-HT₁ autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT_{1A} receptors, to SSRI treatment decouples the feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha1-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha2-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha2-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. Expert Opin Investig Drugs 1999 Nov;8(11):1837-1848

The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis. NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified can be divided into five main categories depending on their chemical structure: indoles, tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats. *Neuroreport* 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a potentially useful biological target for the reduction of brain injury.

Panel 1.3D Summary: Ag2970 Expression of the NOV17c gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2970 Expression of the NOV17c gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV19a

Expression of gene NOV19a was assessed using the primer-probe set Ag2972, described in Table OA. Results of the RTQ-PCR runs are shown in Tables OB, OC, and OD.

Table OA. Probe Name Ag2972

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atgtgtttcagcttccattctg-3'	22	543	1045
Probe	TET-5'-taggtctcggtcattaaccactttt-3'-TAMRA	26	565	1046
Reverse	5'-tgtcctgacacaccaatgatag-3'	22	611	1047

5

Table OB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2972, Run 211008969	Tissue Name	Rel. Exp.(%) Ag2972, Run 211008969
AD 1 Hippo	2.1	Control (Path) 3 Temporal Ctx	1.9
AD 2 Hippo	13.7	Control (Path) 4 Temporal Ctx	70.2
AD 3 Hippo	2.2	AD 1 Occipital Ctx	3.3
AD 4 Hippo	1.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	75.8	AD 3 Occipital Ctx	2.2
AD 6 Hippo	33.4	AD 4 Occipital Ctx	46.3
Control 2 Hippo	11.9	AD 5 Occipital Ctx	22.5
Control 4 Hippo	2.5	AD 6 Occipital Ctx	19.8
Control (Path) 3 Hippo	1.3	Control 1 Occipital Ctx	1.1
AD 1 Temporal Ctx	2.1	Control 2 Occipital Ctx	42.0
AD 2 Temporal Ctx	28.3	Control 3 Occipital Ctx	21.9
AD 3 Temporal Ctx	1.4	Control 4 Occipital Ctx	10.4
AD 4 Temporal Ctx	44.8	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	78.5	Control (Path) 2 Occipital Ctx	3.5
AD 5 Sup Temporal Ctx	12.0	Control (Path) 3 Occipital Ctx	1.1
AD 6 Inf Temporal Ctx	33.7	Control (Path) 4 Occipital Ctx	26.6
AD 6 Sup Temporal Ctx	56.3	Control 1 Parietal Ctx	0.6
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	18.8

Control 2 Temporal Ctx	10.4	Control 3 Parietal Ctx	17.7
Control 3 Temporal Ctx	11.5	Control (Path) 1 Parietal Ctx	77.4
Control 3 Temporal Ctx	7.8	Control (Path) 2 Parietal Ctx	10.3
Control (Path) 1 Temporal Ctx	52.9	Control (Path) 3 Parietal Ctx	0.7
Control (Path) 2 Temporal Ctx	7.9	Control (Path) 4 Parietal Ctx	57.8

Table OC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2972, Run 164314417	Tissue Name	Rel. Exp.(%) Ag2972, Run 164314417
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	11.3
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	4.1	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	2.1	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	12.2	Lung Microvascular EC TNFalpha + IL-1beta	4.2
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	37.4	Bronchial epithelium TNFalpha + IL1beta	7.6
Primary Th2 rest	46.7	Small airway epithelium none	0.0
Primary Tr1 rest	9.7	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	21.2	Coronary artery SMC rest	22.2
CD45RO CD4 lymphocyte act	16.3	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	8.4	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	11.6	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	20.3

CD4 lymphocyte none	25.5	KU-812 (Basophil) PMA/ionomycin	34.6
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	26.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.5
LAK cells IL-2	46.7	Liver cirrhosis	22.8
LAK cells IL-2+IL-12	32.5	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	49.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	38.2	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	5.7	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	7.7	NCI-H292 IL-13	0.0
Two Way MLR 3 day	73.2	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	24.8	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	1.7	Lung fibroblast none	0.0
PBMC PWM	100.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	41.2	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	10.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	40.1	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	21.6	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	6.6
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	21.9	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti- CD40	0.0	IBD Colitis 2	2.9
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	9.2	Colon	7.6
Macrophages rest	19.6	Lung	0.0
Macrophages LPS	8.1	Thymus	11.8
HUVEC none	0.0	Kidney	71.7
HUVEC starved	0.0		

Table OD. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag2972, Run 171670020	Tissue Name	Rel. Exp.(%) Ag2972, Run 171670020
BA4 Control	17.8	BA17 PSP	52.5
BA4 Control2	8.6	BA17 PSP2	3.1
BA4 Alzheimer's2	7.1	Sub Nigra Control	31.0
BA4 Parkinson's	42.9	Sub Nigra Control2	14.7
BA4 Parkinson's2	51.1	Sub Nigra Alzheimer's2	5.1
BA4 Huntington's	7.4	Sub Nigra Parkinson's2	28.9
BA4 Huntington's2	0.0	Sub Nigra Huntington's	100.0
BA4 PSP	0.0	Sub Nigra Huntington's2	11.2
BA4 PSP2	3.7	Sub Nigra PSP2	1.3
BA4 Depression	16.7	Sub Nigra Depression	7.4
BA4 Depression2	0.0	Sub Nigra Depression2	2.5
BA7 Control	15.7	Glob Palladus Control	7.8
BA7 Control2	15.7	Glob Palladus Control2	0.0
BA7 Alzheimer's2	0.0	Glob Palladus Alzheimer's	0.0
BA7 Parkinson's	24.3	Glob Palladus Alzheimer's2	2.0
BA7 Parkinson's2	29.9	Glob Palladus Parkinson's	57.0
BA7 Huntington's	43.5	Glob Palladus Parkinson's2	3.8
BA7 Huntington's2	7.2	Glob Palladus PSP	0.0
BA7 PSP	20.2	Glob Palladus PSP2	3.4
BA7 PSP2	2.1	Glob Palladus Depression	19.5
BA7 Depression	5.2	Temp Pole Control	0.0
BA9 Control	0.0	Temp Pole Control2	3.3
BA9 Control2	22.7	Temp Pole Alzheimer's	0.0
BA9 Alzheimer's	1.3	Temp Pole Alzheimer's2	8.6
BA9	9.0	Temp Pole	0.0

increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; further more the purinergic receptors have also been implicated as drug targets in the treatment of cerebral ischemia. The b-adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the a-adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

References:

- El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. *Br J Pharmacol* 2001 Sep;134(1):68-77
1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg⁻¹), i.p.) and KW 6002 (0.1 - 10 mg kg⁻¹), p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg⁻¹) and ZM 241385 (15 - 60 mg kg⁻¹) were effective in mice previously screened for having high immobility time, while SCH 58261 at 10 mg kg⁻¹ reduced immobility of mice that were selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg⁻¹ reduced the immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg⁻¹ by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg⁻¹) i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg⁻¹) i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants.

Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.

Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. Clin Psychiatry 2001;62 Suppl 15:12-7

5 Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT_{1A} (cell body) and 5-HT_{1B} (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term
10 treatment desensitizes the inhibitory 5-HT₁ autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT_{1A} receptors, to SSRI treatment decouples the feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their
15 connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha₁-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha₂-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha₂-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the
20 NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. Expert Opin Investig
25 Drugs 1999 Nov;8(11):1837-1848

 The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review
30 focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis.

NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified can be divided into five main categories depending on their chemical structure: indoles,
5 tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats. Neuroreport 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global
10 ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were
15 observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a potentially useful biological target for the reduction of brain injury.

Panel 1.3D Summary: Ag2972 Expression of the NOV19a gene is low/undetectable
20 in all samples on this panel ($CTs > 35$). (Data not shown.)

Panel 4D Summary: Ag2972 Expression of the the NOV19a gene is restricted to a few samples in this panel, with highest expression in peripheral blood mononuclear cells (PBMC) treated with the B cell selective pokeweed mitogen. No significant levels of expression of the transcript are seen in PBMC that contain normal B cells. Therefore, the
25 putative GPCR encoded by this gene could potentially be used diagnostically to identify activated B cells. In addition, the gene product could also potentially be used therapeutically in the treatment of diseases in which B cells are activated.

Panel CNS_1 Summary: Ag2972 This panel confirms expression of the NOV19a gene in the brains of an independent group of subjects. Please see panel 1.3d for a discussion
30 of utility of this gene in the central nervous system.

NOV20

Expression of gene NOV20 was assessed using the primer-probe set Ag2973, described in Table PA. Results of the RTQ-PCR runs are shown in Tables PB and PC.

Table PA. Probe Name Ag2973

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gctgtgtggctcaagtctatct-3'	22	287	1048
Probe	TET-5'-ttctctgcctttgcatctgctgagct-3'- TAMRA	26	310	1049
Reverse	5'-agcggtcataagacatgacagt-3'	22	346	1050

Table PB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2973, Run 209779096	Tissue Name	Rel. Exp.(%) Ag2973, Run 209779096
AD 1 Hippo	4.9	Control (Path) 3 Temporal Ctx	1.7
AD 2 Hippo	16.2	Control (Path) 4 Temporal Ctx	46.7
AD 3 Hippo	2.2	AD 1 Occipital Ctx	3.9
AD 4 Hippo	3.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	50.7	AD 3 Occipital Ctx	1.7
AD 6 Hippo	16.4	AD 4 Occipital Ctx	19.1
Control 2 Hippo	8.2	AD 5 Occipital Ctx	24.5
Control 4 Hippo	0.5	AD 6 Occipital Ctx	2.7
Control (Path) 3 Hippo	0.0	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	4.5	Control 2 Occipital Ctx	17.8
AD 2 Temporal Ctx	27.7	Control 3 Occipital Ctx	8.7
AD 3 Temporal Ctx	3.6	Control 4 Occipital Ctx	0.7
AD 4 Temporal Ctx	17.6	Control (Path) 1 Occipital Ctx	74.7
AD 5 Inf Temporal Ctx	64.2	Control (Path) 2 Occipital Ctx	13.2
AD 5 Sup Temporal Ctx	21.8	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	24.3	Control (Path) 4 Occipital Ctx	9.2
AD 6 Sup Temporal Ctx	22.2	Control 1 Parietal Ctx	1.5
Control 1 Temporal Ctx	1.4	Control 2 Parietal Ctx	27.9
Control 2 Temporal Ctx	33.9	Control 3 Parietal Ctx	16.8
Control 3 Temporal	18.4	Control (Path) 1	95.3

Ctx		Parietal Ctx	
Control 3 Temporal Ctx	3.2	Control (Path) 2 Parietal Ctx	23.7
Control (Path) 1 Temporal Ctx	100.0	Control (Path) 3 Parietal Ctx	1.4
Control (Path) 2 Temporal Ctx	68.8	Control (Path) 4 Parietal Ctx	33.7

Table PC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2973, Run 164329850	Tissue Name	Rel. Exp.(%) Ag2973, Run 164329850
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	11.7	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	6.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-	0.0	CCD1106 (Keratinocytes)	0.0

receptors contains a large number of neurotransmitter receptors, including the dopamine, serotonin, α and β -adrenergic, acetylcholine muscarinic, histamine, peptide, and metabotropic glutamate receptors. GPCRs are excellent drug targets in various neurologic and psychiatric diseases. All antipsychotics have been shown to act at the dopamine D2 receptor; similarly novel antipsychotics also act at the serotonergic receptor, and often the muscarinic and adrenergic receptors as well. While the majority of antidepressants can be classified as selective serotonin reuptake inhibitors, blockade of the 5-HT1A and α_2 adrenergic receptors increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; furthermore the purinergic receptors have also been implicated as drug targets in the treatment of cerebral ischemia. The β -adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the α -adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

References:

El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. *Br J Pharmacol* 2001 Sep;134(1):68-77

1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg⁻¹, i.p.) and KW 6002 (0.1 - 10 mg kg⁻¹, p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg⁻¹) and ZM 241385 (15 - 60 mg kg⁻¹) were effective in mice previously screened for having high immobility time, while SCH 58261 at 10 mg kg⁻¹ reduced immobility of mice that were selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg⁻¹ reduced the

immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg(-1) by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg(-1) i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg(-1) i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants. Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.

Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. Clin Psychiatry 2001;62 Suppl 15:12-7

Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT1A (cell body) and 5-HT1B (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term treatment desensitizes the inhibitory 5-HT1 autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT1A receptors, to SSRI treatment decouples the feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha1-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha2-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha2-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. Expert Opin Investig Drugs 1999 Nov;8(11):1837-1848

The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the

metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis. NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified can be divided into five main categories depending on their chemical structure: indoles, tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats. *Neuroreport* 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a potentially useful biological target for the reduction of brain injury.

Panel 1.3D Summary: Ag2973 Expression of the NOV20 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 3D Summary: Ag2973 Expression of the NOV20 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2973 Significant expression of the NOV20 gene is detected in a liver cirrhosis sample (CT = 32.7). Furthermore, expression of this gene is not detected in normal liver in Panel 1.3D, suggesting that its expression is unique to liver cirrhosis. This gene encodes a putative GPCR; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative GPCR could also be used for the diagnosis of liver cirrhosis.

Panel CNS_1 Summary: Ag2973 Expression of the NOV20 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV16b

Expression of gene NOV16b was assessed using the primer-probe sets Ag2875 and Ag3010, described in Tables QA and QB. Results of the RTQ-PCR runs are shown in Table QC.

Table QA. Probe Name Ag2875

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaacatcatctcctaccctgaa-3'	22	267	1051
Probe	TET-5'-tgcatgactcagctctacttcttcctcg-3'-TAMRA	28	289	1052
Reverse	5'-atgtgacactctgcaatagcaa-3'	22	320	1053

Table QB. Probe Name Ag3010

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggggaaagtatcctctgtgttt-3'	22	810	1054
Probe	TET-5'-ttattgtgcccattgtgaaccctctg-3'-TAMRA	26	839	1055
Reverse	5'-cagggaaacatggacatcttta-3'	22	882	1056

10

Table QC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2875, Run 164311029	Tissue Name	Rel. Exp.(%) Ag2875, Run 164311029
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	3.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0

Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	24.0
2ry Th1/Th2/Tr1 _anti- CD95 CH11	5.1	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	72.2
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	20.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	1.4
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	3.2
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	3.1	Dermal fibroblast	0.0

		CCD1070 TNF alpha	
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	33.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	2.4	Dermal fibroblast IL-4	1.9
Dendritic cells anti- CD40	100.0	IBD Colitis 2	16.3
Monocytes rest	0.0	IBD Crohn's	6.7
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	44.1	Lung	21.8
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	3.1	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2875/Ag3010 Results from two experiments with the NOV16b gene are not included. The amp plots indicate that there were experimental difficulties with these runs.

5 **Panel 2.2 Summary:** Ag2875 Expression of the NOV16b gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2875 Highest expression of the NOV15a is in anti-CD40 treated dendritic cells (CT=33.2), with much lower expression in untreated dendritic cells. Thus, this gene product may be important in dendritic cell activation. Significant expression of this gene is also seen in liver cirrhosis. This gene encodes a putative GPCR; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative GPCR could also be used for the diagnosis of liver cirrhosis. In addition, significant expression of this gene is seen in resting macrophages. The putative GPCR encoded for by this transcript may therefore be important in macrophage detection of chemokine gradients and trafficking into specific sites within a tissue and in activation. Antibody or protein therapeutics designed against the protein encoded for by this transcript could reduce or inhibit inflammation in asthma, emphysema, allergy, psoriasis, arthritis, or any other condition in which macrophage localization/activation is important. A second experiment with the probe/primer set Ag3010 shows low/undetectable expression in all samples on this panel (CTs>35). (Data not shown.)

10

15

20

NOV21a

Expression of gene NOV21a was assessed using the primer-probe sets Ag2963 and Ag1292, described in Tables RA and RB. Results of the RTQ-PCR runs are shown in Tables RC, RD and RE.

Table RA. Probe Name Ag2963

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggtagccaccctttgttccaat-3'	21	903	1057
Probe	TET-5'-aaacacagccctccaccctagct-3'-TAMRA	24	924	1058
Reverse	5'-gtcctcgctgtgacactga-3'	19	963	1059

5

Table RB. Probe Name Ag1292

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cagctgtgagaggctggata-3'	20	731	1060
Probe	TET-5'-atccaggagtgaccaccacgtgact-3'-TAMRA	25	760	1061
Reverse	5'-tcctgttgctttcacgtagagt-3'	22	797	1062

Table RC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2963, Run 209778824	Rel. Exp.(%) Ag2963, Run 230512509	Tissue Name	Rel. Exp.(%) Ag2963, Run 209778824	Rel. Exp.(%) Ag2963, Run 230512509
AD 1 Hippo	6.5	6.8	Control (Path) 3 Temporal Ctx	6.1	3.6
AD 2 Hippo	21.2	16.8	Control (Path) 4 Temporal Ctx	32.5	26.2
AD 3 Hippo	5.5	6.3	AD 1 Occipital Ctx	5.3	4.9
AD 4 Hippo	7.4	7.9	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	21.0	22.2	AD 3 Occipital Ctx	7.9	6.2
AD 6 Hippo	12.2	19.6	AD 4 Occipital Ctx	12.1	13.8
Control 2	58.6	44.8	AD 5	6.7	3.3

Hippo			Occipital Ctx		
Control 4 Hippo	12.9	12.9	AD 6 Occipital Ctx	73.7	81.8
Control (Path) 3 Hippo	2.0	4.5	Control 1 Occipital Ctx	5.9	2.5
AD 1 Temporal Ctx	2.3	6.3	Control 2 Occipital Ctx	89.5	100.0
AD 2 Temporal Ctx	13.2	19.5	Control 3 Occipital Ctx	27.4	25.7
AD 3 Temporal Ctx	3.3	7.1	Control 4 Occipital Ctx	11.3	12.4
AD 4 Temporal Ctx	15.8	20.7	Control (Path) 1 Occipital Ctx	100.0	93.3
AD 5 Inf Temporal Ctx	16.7	21.5	Control (Path) 2 Occipital Ctx	8.4	11.0
AD 5 Sup Temporal Ctx	12.5	11.8	Control (Path) 3 Occipital Ctx	4.4	2.2
AD 6 Inf Temporal Ctx	25.0	25.2	Control (Path) 4 Occipital Ctx	16.7	20.6
AD 6 Sup Temporal Ctx	22.7	17.1	Control 1 Parietal Ctx	9.8	4.6
Control 1 Temporal Ctx	2.1	6.1	Control 2 Parietal Ctx	8.2	8.4
Control 2 Temporal Ctx	82.4	51.8	Control 3 Parietal Ctx	26.6	24.7
Control 3 Temporal Ctx	31.2	20.2	Control (Path) 1 Parietal Ctx	94.6	66.9
Control 4 Temporal Ctx	11.6	12.9	Control (Path) 2 Parietal Ctx	19.3	24.8
Control (Path) 1 Temporal Ctx	77.9	56.3	Control (Path) 3 Parietal Ctx	2.6	3.8

Control (Path) 2 Temporal Ctx	17.0	12.3	Control (Path) 4 Parietal Ctx	35.8	23.0
----------------------------------	------	------	-------------------------------------	------	------

Table RD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2963, Run 167809191	Tissue Name	Rel. Exp.(%) Ag2963, Run 167809191
Liver adenocarcinoma	82.4	Kidney (fetal)	45.4
Pancreas	50.0	Renal ca. 786-0	53.6
Pancreatic ca. CAPAN 2	10.7	Renal ca. A498	19.6
Adrenal gland	14.2	Renal ca. RXF 393	21.2
Thyroid	26.2	Renal ca. ACHN	12.5
Salivary gland	21.2	Renal ca. UO-31	13.7
Pituitary gland	31.6	Renal ca. TK-10	27.2
Brain (fetal)	40.9	Liver	49.7
Brain (whole)	71.7	Liver (fetal)	45.1
Brain (amygdala)	11.6	Liver ca. (hepatoblast) HepG2	16.3
Brain (cerebellum)	62.4	Lung	26.4
Brain (hippocampus)	47.0	Lung (fetal)	66.0
Brain (substantia nigra)	89.5	Lung ca. (small cell) LX-1	42.0
Brain (thalamus)	53.2	Lung ca. (small cell) NCI-H69	21.6
Cerebral Cortex	92.7	Lung ca. (s.cell var.) SHP-77	59.9
Spinal cord	30.6	Lung ca. (large cell)NCI-H460	12.9
glio/astro U87-MG	28.5	Lung ca. (non-sm. cell) A549	37.1
glio/astro U-118-MG	44.4	Lung ca. (non-s.cell) NCI-H23	47.3
astrocytoma SW1783	50.3	Lung ca. (non-s.cell) HOP-62	11.8
neuro*; met SK-N-AS	40.6	Lung ca. (non-s.cl) NCI-H522	16.8
astrocytoma SF-539	37.1	Lung ca. (squam.) SW 900	29.3
astrocytoma SNB-75	49.3	Lung ca. (squam.) NCI-H596	51.1
glioma SNB-19	31.0	Mammary gland	70.2
glioma U251	52.9	Breast ca.* (pl.ef) MCF-7	15.2

glioma SF-295	27.0	Breast ca.* (pl.ef) MDA-MB-231	29.7
Heart (fetal)	90.8	Breast ca.* (pl.ef) T47D	41.2
Heart	22.7	Breast ca. BT-549	28.3
Skeletal muscle (fetal)	83.5	Breast ca. MDA-N	54.3
Skeletal muscle	57.0	Ovary	47.0
Bone marrow	25.2	Ovarian ca. OVCAR-3	26.1
Thymus	74.7	Ovarian ca. OVCAR-4	18.7
Spleen	28.1	Ovarian ca. OVCAR-5	82.9
Lymph node	54.7	Ovarian ca. OVCAR-8	3.4
Colorectal	7.9	Ovarian ca. IGROV-1	22.2
Stomach	7.0	Ovarian ca.* (ascites) SK-OV-3	100.0
Small intestine	25.3	Uterus	28.3
Colon ca. SW480	25.9	Placenta	3.2
Colon ca.* SW620(SW480 met)	40.1	Prostate	22.1
Colon ca. HT29	24.1	Prostate ca.* (bone met)PC-3	29.3
Colon ca. HCT-116	35.8	Testis	19.3
Colon ca. CaCo-2	52.9	Melanoma Hs688(A).T	19.5
Colon ca. tissue(ODO3866)	4.4	Melanoma* (met) Hs688(B).T	22.4
Colon ca. HCC-2998	31.9	Melanoma UACC-62	42.9
Gastric ca.* (liver met) NCI-N87	27.2	Melanoma M14	8.3
Bladder	20.6	Melanoma LOX IMVI	25.0
Trachea	11.7	Melanoma* (met) SK-MEL-5	23.3
Kidney	56.6	Adipose	21.0

Table RE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1292, Run	Rel. Exp.(%) Ag2963, Run	Tissue Name	Rel. Exp.(%) Ag1292, Run	Rel. Exp.(%) Ag2963, Run
-------------	-----------------------------------	-----------------------------------	-------------	-----------------------------------	-----------------------------------

	138719232	164333805		138719232	164333805
Secondary Th1 act	14.5	18.6	HUVEC IL-1beta	7.6	3.0
Secondary Th2 act	11.3	47.0	HUVEC IFN gamma	13.5	31.6
Secondary Tr1 act	23.2	29.1	HUVEC TNF alpha + IFN gamma	2.2	26.1
Secondary Th1 rest	4.0	8.8	HUVEC TNF alpha + IL4	8.3	33.2
Secondary Th2 rest	5.0	18.2	HUVEC IL-11	2.8	26.6
Secondary Tr1 rest	5.8	2.2	Lung Microvascular EC none	7.1	62.0
Primary Th1 act	19.1	55.9	Lung Microvascular EC TNFalpha + IL- 1beta	2.7	40.9
Primary Th2 act	16.4	33.4	Microvascular Dermal EC none	8.8	41.2
Primary Tr1 act	25.3	66.9	Microvascular Dermal EC TNFalpha + IL- 1beta	7.5	31.6
Primary Th1 rest	13.6	33.0	Bronchial epithelium TNFalpha + IL1beta	8.7	37.1
Primary Th2 rest	9.3	16.7	Small airway epithelium none	1.7	8.4
Primary Tr1 rest	18.8	23.0	Small airway epithelium TNFalpha + IL- 1beta	6.6	14.2
CD45RA CD4 lymphocyte act	7.9	39.0	Coronary artery SMC rest	55.9	21.6
CD45RO CD4 lymphocyte act	10.3	27.2	Coronary artery SMC TNFalpha + IL-1beta	5.3	14.6
CD8 lymphocyte act	9.9	29.7	Astrocytes rest	5.8	24.1
Secondary CD8 lymphocyte rest	11.9	44.4	Astrocytes TNFalpha + IL- 1beta	4.4	19.5
Secondary CD8 lymphocyte act	13.0	22.8	KU-812 (Basophil) rest	16.0	66.9
CD4 lymphocyte none	6.3	21.3	KU-812 (Basophil)	8.1	87.7

			PMA/ionomycin		
2ry Th1/Th2/Tr1_anti- CD95 CH11	7.9	20.3	CCD1106 (Keratinocytes) none	5.8	41.2
LAK cells rest	10.1	39.5	CCD1106 (Keratinocytes) TNFalpha + IL- 1beta	6.5	11.5
LAK cells IL-2	5.7	33.2	Liver cirrhosis	4.9	7.9
LAK cells IL-2+IL- 12	9.0	34.2	Lupus kidney	4.5	9.2
LAK cells IL- 2+IFN gamma	12.4	31.9	NCI-H292 none	11.5	63.7
LAK cells IL-2+ IL-18	16.8	16.2	NCI-H292 IL-4	17.7	77.9
LAK cells PMA/ionomycin	15.4	50.7	NCI-H292 IL-9	59.5	54.3
NK Cells IL-2 rest	7.6	22.8	NCI-H292 IL-13	44.1	63.7
Two Way MLR 3 day	7.7	49.0	NCI-H292 IFN gamma	10.4	89.5
Two Way MLR 5 day	5.4	29.9	HPAEC none	8.2	28.1
Two Way MLR 7 day	5.3	17.0	HPAEC TNF alpha + IL-1 beta	9.6	30.1
PBMC rest	5.2	11.2	Lung fibroblast none	8.7	28.7
PBMC PWM	17.1	100.0	Lung fibroblast TNF alpha + IL-1 beta	4.5	6.9
PBMC PHA-L	10.9	51.4	Lung fibroblast IL-4	10.2	53.6
Ramos (B cell) none	24.1	88.3	Lung fibroblast IL-9	9.7	49.7
Ramos (B cell) ionomycin	14.6	92.7	Lung fibroblast IL-13	22.1	44.1
B lymphocytes PWM	11.8	59.0	Lung fibroblast IFN gamma	9.5	29.5
B lymphocytes CD40L and IL-4	20.7	21.8	Dermal fibroblast CCD1070 rest	32.3	37.6
EOL-1 dbcAMP	23.3	32.8	Dermal fibroblast CCD1070 TNF alpha	100.0	39.0
EOL-1 dbcAMP PMA/ionomycin	4.2	25.5	Dermal fibroblast CCD1070 IL-1 beta	14.3	28.7
Dendritic cells	4.3	22.8	Dermal fibroblast	7.5	17.8

none			IFN gamma		
Dendritic cells LPS	4.4	10.5	Dermal fibroblast IL-4	30.4	48.6
Dendritic cells anti-CD40	6.4	24.3	IBD Colitis 2	1.3	1.2
Monocytes rest	7.6	38.4	IBD Crohn's	2.6	2.2
Monocytes LPS	4.5	8.0	Colon	7.3	37.9
Macrophages rest	6.3	21.5	Lung	5.8	22.8
Macrophages LPS	5.5	1.9	Thymus	12.7	39.2
HUVEC none	4.8	30.8	Kidney	19.1	98.6
HUVEC starved	7.8	27.7			

CNS_neurodegeneration_v1.0 Summary: Ag2963 The NOV21a gene, a secretory protease homolog, appears to be downregulated in the temporal cortex of Alzheimer's disease patients when compared to non-demented controls. Up regulation of this protease may therefore be of use in the treatment of Alzheimer's, particularly because Alzheimer's disease is believed to result at least in part from the improper processing of proteins (APP, Tau). This protease may serve to lower the levels of these disease proteins and ameliorate the dementia/pathology associated with Alzheimer's.

Panel 1.3D Summary: Ag2963 The NOV21a gene, a putative secretory serine-protease, is widely expressed in this panel. Highest expression is in an ovarian cancer cell line (CT=32), with expression detected in all cancer cell lines in this panel. Thus, inhibition of the protease domain might lead to a decrease in cell survival and proliferation and serve as a small molecule target in cancer.

This gene product also has low levels of expression in pancreas, thyroid, pituitary, adipose, and adult and fetal types of heart, skeletal muscle and liver. Therefore, this serine protease-like gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

The expression in this panel further confirms expression of this gene in the CNS. Please see CNS_neurodegeneration for discussion of utility of this gene in the CNS.

Panel 4D Summary: Ag1292/Ag2693 The NOV21a transcript is expressed on most tissues in panel 4D. This widespread expression is consistent with the results in Panel 1.3D. This transcript encodes a serine protease like protein with potential enzymatic activity and may important in maintaining normal cellular functions in a number of tissues. Therefore, therapies designed with the protein encoded by this transcript could be important in regulating cellular viability or function.

NOV22a

Expression of gene NOV22a was assessed using the primer-probe set Ag2964, described in Table SA.

Table SA. Probe Name Ag2964

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaaatacgcaggaaggaaatct-3'	22	473	1063
Probe	TET-5'-cgtgggcatcatagaccagaaaacct-3'-TAMRA	26	507	1064
Reverse	5'-ggtctgtgagggagaagtgtga-3'	22	544	1065

CNS_neurodegeneration_v1.0 Summary: Ag2964 Expression of the NOV22a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 1.3D Summary: Ag2964 Expression of the NOV22a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 4D Summary: Ag2964 Expression of the NOV22a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

NOV23a

Expression of gene NOV23a was assessed using the primer-probe set Ag2967, described in Table TA.

Table TA. Probe Name Ag2967

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctcaaggaggacttccgtaag-3'	21	478	1066
Probe	TET-5'-atcttcctctgcagaaagccactgtg-3'-TAMRA	26	500	1067
Reverse	5'-agtgagtgaatggccgtaca-3'	20	557	1068

CNS_neurodegeneration_v1.0 Summary: Ag2967 Expression of the NOV23a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 1.3D Summary: Ag2967 Expression of the NOV23a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 4D Summary: Ag2967 Expression of the NOV23a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

NOV23a and NOV23b

Expression of gene NOV23a and variant NOV23b was assessed using the primer-probe set Ag2996, described in Table UA. Results of the RTQ-PCR runs are shown in Table UB.

10

Table UA. Probe Name Ag2996

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cagagggttacgccctgaagt-3'	20	54	1069
Probe	TET-5'-ctcagatcctgggccaggcactg-3'-TAMRA	23	77	1070
Reverse	5'-cagaaagaggtcctgctcatg-3'	21	113	1071

Table UB. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2996, Run 165296353	Tissue Name	Rel. Exp.(%) Ag2996, Run 165296353
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	23.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium	0.0

		TNFalpha + IL-1beta	
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	11.3	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	14.6	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	15.3	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0

Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	49.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2996 Results from one experiment with the NOV23a gene are not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 4D Summary: Ag2996 Significant expression of the NOV23a gene is detected in a liver cirrhosis sample and normal lung tissue(CTs=33-35). Thus, antibodies to this protein product could potentially be used for the diagnosis of liver cirrhosis or as a marker of normal lung tissue. Furthermore, therapeutic modulation of the expression or function of this gene may reduce or inhibit fibrosis that occurs in liver cirrhosis.

NOV24a

Expression of gene NOV24a was assessed using the primer-probe set Ag2934, described in Table VA.

Table VA. Probe Name Ag2934

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaatctaattccgcaggaaatg-3'	22	1645	1072
Probe	TET-5'-cgtggttgccaaattctcagcgttta-3'-TAMRA	26	1668	1073
Reverse	5'-tcttttcatgaacctcatttgc-3'	22	1714	1074

CNS_neurodegeneration_v1.0 Summary: Ag2934 Expression of the NOV24a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 1.3D Summary: Ag2934 Expression of the NOV24a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2934 Expression of the NOV24a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

NOV25

Expression of gene NOV25 was assessed using the primer-probe sets Ag2935 and Ag3039, described in Tables WA and WB. Results of the RTQ-PCR runs are shown in Tables WC, WD and WE.

Table WA. Probe Name Ag2935

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctggactctacctcggaaactt-3'	22	88	1075
Probe	TET-5'-ctgggccgaaataagatcacacacat-3'-TAMRA	26	135	1076
Reverse	5'-gggtgactcatggatagagatg-3'	22	161	1077

5

Table WB. Probe Name Ag3039

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gccgaaataagatcacacacat-3'	22	139	1078
Probe	TET-5'-tctatccatgagtcaccccagcctct-3'-TAMRA	26	165	1079
Reverse	5'-atgcgaaggtaggtgatatcct-3'	22	196	1080

Table WC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2935, Run 209777519	Rel. Exp.(%) Ag3039, Run 211012103	Tissue Name	Rel. Exp.(%) Ag2935, Run 209777519	Rel. Exp.(%) Ag3039, Run 211012103
AD 1 Hippo	34.2	18.4	Control (Path) 3 Temporal Ctx	11.9	8.2
AD 2 Hippo	59.5	48.0	Control (Path) 4 Temporal Ctx	33.7	36.3
AD 3 Hippo	15.3	9.8	AD 1 Occipital Ctx	24.3	9.5
AD 4 Hippo	18.7	13.6	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 Hippo	90.8	70.2	AD 3 Occipital Ctx	12.9	6.3
AD 6 Hippo	50.7	69.3	AD 4 Occipital Ctx	32.8	20.9
Control 2	36.3	25.5	AD 5	66.9	18.3

Hippo			Occipital Ctx		
Control 4 Hippo	33.4	24.0	AD 6 Occipital Ctx	28.9	43.2
Control (Path) 3 Hippo	7.2	7.6	Control 1 Occipital Ctx	7.6	6.0
AD 1 Temporal Ctx	20.7	24.3	Control 2 Occipital Ctx	90.8	57.0
AD 2 Temporal Ctx	53.2	36.9	Control 3 Occipital Ctx	24.7	18.7
AD 3 Temporal Ctx	5.8	4.7	Control 4 Occipital Ctx	17.7	13.9
AD 4 Temporal Ctx	35.8	24.5	Control (Path) 1 Occipital Ctx	93.3	74.2
AD 5 Inf Temporal Ctx	100.0	100.0	Control (Path) 2 Occipital Ctx	19.1	14.8
AD 5 Sup Temporal Ctx	81.2	62.9	Control (Path) 3 Occipital Ctx	8.6	4.3
AD 6 Inf Temporal Ctx	62.4	58.2	Control (Path) 4 Occipital Ctx	27.9	25.2
AD 6 Sup Temporal Ctx	50.3	49.3	Control 1 Parietal Ctx	16.2	15.9
Control 1 Temporal Ctx	14.5	11.6	Control 2 Parietal Ctx	68.3	58.2
Control 2 Temporal Ctx	51.4	34.4	Control 3 Parietal Ctx	31.0	32.1
Control 3 Temporal Ctx	33.0	20.0	Control (Path) 1 Parietal Ctx	69.7	66.9
Control 3 Temporal Ctx	19.6	20.7	Control (Path) 2 Parietal Ctx	31.2	39.0

Control (Path) 1 Temporal Ctx	55.1	44.4	Control (Path) 3 Parietal Ctx	3.1	4.6
Control (Path) 2 Temporal Ctx	56.3	30.4	Control (Path) 4 Parietal Ctx	61.6	35.8

Table WD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2935, Run 167646849	Rel. Exp.(%) Ag3039, Run 167961816	Tissue Name	Rel. Exp.(%) Ag2935, Run 167646849	Rel. Exp.(%) Ag3039, Run 167961816
Liver adenocarcinoma	1.4	1.7	Kidney (fetal)	49.3	38.2
Pancreas	1.2	2.0	Renal ca. 786-0	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	1.7	0.4
Adrenal gland	3.1	1.2	Renal ca. RXF 393	0.0	0.1
Thyroid	11.7	6.0	Renal ca. ACHN	5.7	1.5
Salivary gland	1.3	0.8	Renal ca. UO-31	0.0	0.0
Pituitary gland	6.3	3.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	10.0	7.4	Liver	0.7	0.3
Brain (whole)	12.8	7.7	Liver (fetal)	1.8	0.7
Brain (amygdala)	6.9	6.0	Liver ca. (hepatoblast) HepG2	1.0	0.1
Brain (cerebellum)	9.3	5.9	Lung	1.3	1.3
Brain (hippocampus)	12.1	3.9	Lung (fetal)	6.5	3.0
Brain (substantia nigra)	24.0	16.0	Lung ca. (small cell) LX-1	0.3	0.0
Brain (thalamus)	5.1	3.8	Lung ca. (small cell) NCI-H69	2.8	1.5
Cerebral Cortex	15.6	10.8	Lung ca. (s.cell var.) SHP-77	7.2	3.7
Spinal cord	37.4	15.2	Lung ca. (large	0.0	0.1

			cell)NCI-H460		
glio/astro U87-MG	0.0	0.2	Lung ca. (non-sm. cell) A549	2.4	1.6
glio/astro U-118-MG	0.4	0.1	Lung ca. (non-s.cell) NCI-H23	4.0	0.4
astrocytoma SW1783	3.5	0.6	Lung ca. (non-s.cell) HOP-62	1.1	0.0
neuro*; met SK-N-AS	0.8	0.3	Lung ca. (non-s.cl) NCI-H522	3.9	1.4
astrocytoma SF-539	3.0	1.5	Lung ca. (squam.) SW 900	2.7	0.7
astrocytoma SNB-75	2.5	1.2	Lung ca. (squam.) NCI-H596	12.2	5.7
glioma SNB-19	2.7	2.1	Mammary gland	3.9	2.4
glioma U251	1.0	0.3	Breast ca.* (pl.ef) MCF-7	0.7	0.7
glioma SF-295	2.9	2.3	Breast ca.* (pl.ef) MDA-MB-231	0.4	0.1
Heart (fetal)	24.0	14.5	Breast ca.* (pl.ef) T47D	23.3	13.5
Heart	5.6	3.4	Breast ca. BT-549	0.5	1.2
Skeletal muscle (fetal)	11.3	5.1	Breast ca. MDA-N	6.2	6.4
Skeletal muscle	1.1	0.0	Ovary	11.4	3.9
Bone marrow	0.0	0.4	Ovarian ca. OVCAR-3	1.0	0.3
Thymus	1.4	0.2	Ovarian ca. OVCAR-4	21.3	13.8
Spleen	3.5	3.1	Ovarian ca. OVCAR-5	5.6	1.5
Lymph node	0.7	0.9	Ovarian ca. OVCAR-8	0.1	0.0
Colorectal	1.5	0.4	Ovarian ca. IGROV-1	0.0	0.0
Stomach	2.5	0.6	Ovarian ca.* (ascites) SK-OV-3	0.4	0.4
Small intestine	3.4	1.0	Uterus	3.0	2.2
Colon ca. SW480	8.8	6.7	Placenta	0.6	0.0

Colon ca.* SW620(SW480 met)	0.0	0.0	Prostate	3.1	2.7
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC- 3	0.0	0.0
Colon ca. HCT- 116	0.0	0.0	Testis	100.0	100.0
Colon ca. CaCo-2	23.0	12.2	Melanoma Hs688(A).T	0.0	0.1
Colon ca. tissue(ODO3866)	1.3	1.2	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC- 2998	6.4	3.0	Melanoma UACC-62	12.0	5.2
Gastric ca.* (liver met) NCI-N87	2.7	1.2	Melanoma M14	2.7	0.8
Bladder	3.1	1.8	Melanoma LOX IMVI	0.0	0.0
Trachea	1.4	0.6	Melanoma* (met) SK- MEL-5	6.5	2.0
Kidney	84.7	41.8	Adipose	5.0	0.6

Table WE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2935, Run 164403313	Rel. Exp.(%) Ag3039, Run 162427949	Tissue Name	Rel. Exp.(%) Ag2935, Run 164403313	Rel. Exp.(%) Ag3039, Run 162427949
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.2	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.2	0.3
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.3
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	1.3
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	0.0	0.0

Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.3	0.3
Primary Tr1 act	0.0	0.0	Microsvascular Dermal EC TNFalpha + IL- 1beta	0.3	0.4
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	1.7	1.4
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.2	1.1
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL- 1beta	0.3	0.3
CD45RA CD4 lymphocyte act	0.4	0.0	Coronary artery SMC rest	0.2	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.1	0.0	Astrocytes rest	1.9	4.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL- 1beta	2.4	2.3
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.8	0.4
LAK cells rest	0.6	0.1	CCD1106 (Keratinocytes) TNFalpha + IL- 1beta	0.5	0.4
LAK cells IL-2	0.0	0.4	Liver cirrhosis	0.0	0.7
LAK cells IL-2+IL- 12	0.3	0.0	Lupus kidney	4.0	5.2
LAK cells IL- 2+IFN gamma	0.0	0.4	NCI-H292 none	1.4	3.5
LAK cells IL-2+ IL-18	0.8	0.8	NCI-H292 IL-4	1.2	0.9
LAK cells PMA/ionomycin	0.3	0.0	NCI-H292 IL-9	1.6	1.8

NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	2.1	1.2
Two Way MLR 3 day	0.0	0.0	NCI-H292 IFN gamma	2.3	3.6
Two Way MLR 5 day	0.0	0.0	HPAEC none	0.0	0.0
Two Way MLR 7 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	0.0	0.3	Lung fibroblast none	0.0	0.0
PBMC PWM	0.2	2.5	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0
PBMC PHA-L	0.4	1.6	Lung fibroblast IL-4	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IL-13	0.0	0.5
B lymphocytes PWM	3.0	5.5	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes CD40L and IL-4	1.0	1.3	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	0.5	0.6	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.3	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells LPS	0.1	0.4	Dermal fibroblast IL-4	0.0	0.0
Dendritic cells anti-CD40	0.3	0.0	IBD Colitis 2	0.0	0.0
Monocytes rest	0.0	0.0	IBD Crohn's	0.2	1.1
Monocytes LPS	0.0	0.4	Colon	2.1	2.5
Macrophages rest	1.1	1.0	Lung	2.1	5.1
Macrophages LPS	0.3	0.4	Thymus	100.0	100.0
HUVEC none	0.0	0.6	Kidney	3.7	3.1
HUVEC starved	0.0	0.0			

CNS_neurodegeneration_v1.0 Summary: Ag2935/Ag3039 No differential expression of the NOV25 gene is detected between the postmortem brains of Alzheimer's diseased patients and those of non-demented controls. However, this panel confirms the

expression of this gene in the CNS. Please see panel 1.3D for a discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag2935/Ag3039 The expression of the NOV25 gene was assessed in two independent runs with good concordance between runs. Highest expression is seen in the testis (CTs=29). In addition, expression of this gene is extremely low in renal and brain cancer cell lines but is expressed in the normal brain and kidney tissues on this sample. Therefore, this gene may be used as a diagnostic marker for brain and kidney cancer and prostate tissue. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of brain and renal cancers.

This gene encodes a novel protein phosphatase expressed at moderate to low levels in the CNS that may therefore be a small molecule target for the treatment of neurologic diseases.

In addition, this gene is expressed at low levels in metabolic tissues including pancreas, adrenal, thyroid, pituitary, adult and fetal heart, and adipose. This novel protein phosphatase may be a small molecule target for the treatment of metabolic and endocrine disease, including obesity and Types 1 and 2 diabetes. This gene is also differentially expressed in fetal skeletal muscle (CT values = 32-33) when compared to expression in adult skeletal muscle (CT values = 35-40). Therefore, expression of this gene may also be useful for the differentiation of adult and fetal skeletal muscle.

Panel 4D Summary: Ag2935/Ag3039 Expression of the NOV25 gene is highest and almost exclusive to the thymus (CTs=29-30). Expression of this gene could be used to distinguish thymus from the other samples on this panel. The putative phosphatase encoded by this gene may play an important role in T cell development. Small molecule therapeutics designed against the protein encoded by this gene could therefore be utilized to modulate immune function (T cell development) and be important for organ transplant, AIDS treatment or post chemotherapy immune reconstitution.

NOV26a and NOV26b

Expression of gene NOV26a and variant NOV26b was assessed using the primer-probe set Ag2936, described in Table XA.

Table XA. Probe Name Ag2936

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gttctctgctgtctggactttt-3'	21	1	1081

Probe	TET-5'-cccactgagacgcagctgtattctgt-3'- TAMRA	26	27	1082
Reverse	5'-tcgccaatcatatttcacact-3'	22	57	1083

CNS_neurodegeneration_v1.0 Summary: Ag2936 Expression of the NOV26a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

5 **Panel 1.3D Summary:** Ag2936 Expression of the NOV26a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

10 **Panel 4D Summary:** Ag2936 Expression of the NOV26a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

NOV24a and NOV24b

Expression of gene NOV24a and variant NOV24b was assessed using the primer-probe set Ag2966, described in Table YA. Results of the RTQ-PCR runs are shown in Tables YB and YC.

15

Table YA. Probe Name Ag2966

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agcgggcaactcttcatcta-3'	20	1356	1084
Probe	TET-5'-atgcagtcagtgaccagctccctg-3'- TAMRA	24	1377	1085
Reverse	5'-caggacaaagactgcagtcact-3'	22	1409	1086

Table YB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2966, Run 160658385	Rel. Exp.(%) Ag2966, Run 165701959	Tissue Name	Rel. Exp.(%) Ag2966, Run 160658385	Rel. Exp.(%) Ag2966, Run 165701959
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	1.0	0.0
Pancreas	0.2	1.3	Renal ca. 786- 0	0.2	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.0	0.3
Adrenal gland	0.2	0.0	Renal ca. RXF 393	0.0	0.5
Thyroid	0.3	0.0	Renal ca. ACHN	0.3	0.0

Salivary gland	0.1	0.4	Renal ca. UO-31	0.0	0.0
Pituitary gland	0.0	0.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	0.0	0.0	Liver	0.6	0.9
Brain (whole)	0.0	0.3	Liver (fetal)	0.4	0.0
Brain (amygdala)	0.0	0.0	Liver ca. (hepatoblast) HepG2	0.3	0.0
Brain (cerebellum)	0.0	0.0	Lung	0.8	0.6
Brain (hippocampus)	0.5	0.3	Lung (fetal)	1.4	1.4
Brain (substantia nigra)	0.0	0.0	Lung ca. (small cell) LX-1	0.6	0.0
Brain (thalamus)	0.0	0.5	Lung ca. (small cell) NCI-H69	0.0	0.0
Cerebral Cortex	0.0	0.0	Lung ca. (s.cell var.) SHP-77	0.3	0.4
Spinal cord	0.2	0.2	Lung ca. (large cell) NCI-H460	0.2	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non-sm. cell) A549	0.8	0.7
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	1.6	2.2
neuro*; met SK-N-AS	0.0	0.4	Lung ca. (non-s.cl) NCI-H522	0.5	0.0
astrocytoma SF-539	0.0	0.0	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Lung ca. (squam.) NCI-H596	0.0	0.0
glioma SNB-19	0.0	0.4	Mammary gland	0.3	0.7
glioma U251	0.0	0.4	Breast ca.* (pl.ef) MCF-7	0.2	0.0
glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
Heart (fetal)	0.0	0.0	Breast ca.*	0.0	0.0

			(pl.ef) T47D		
Heart	0.0	0.0	Breast ca. BT-549	0.2	0.0
Skeletal muscle (fetal)	3.9	0.3	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.0	0.0	Ovary	0.5	0.0
Bone marrow	6.9	2.5	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	1.9	0.9	Ovarian ca. OVCAR-4	0.0	0.4
Spleen	4.2	1.0	Ovarian ca. OVCAR-5	0.3	0.0
Lymph node	3.0	5.4	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.5	0.0	Ovarian ca. IGROV-1	0.0	0.0
Stomach	0.9	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0	0.0
Small intestine	1.5	0.5	Uterus	0.0	0.5
Colon ca. SW480	0.8	0.0	Placenta	0.0	0.0
Colon ca.* SW620(SW480 met)	0.0	0.0	Prostate	0.0	0.0
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC-3	0.0	0.0
Colon ca. HCT-116	0.0	0.0	Testis	1.6	0.8
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	0.3	0.3	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC-2998	0.3	0.0	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.2	0.7	Melanoma M14	0.0	0.0
Bladder	0.2	0.3	Melanoma LOX IMVI	0.0	0.0
Trachea	0.6	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	100.0	100.0	Adipose	0.1	0.6

Table YC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2966, Run 160660646	Tissue Name	Rel. Exp.(%) Ag2966, Run 160660646
Secondary Th1 act	0.0	HUVEC IL-1beta	0.3
Secondary Th2 act	0.0	HUVEC IFN gamma	0.9
Secondary Tr1 act	0.7	HUVEC TNF alpha + IFN gamma	0.3
Secondary Th1 rest	0.2	HUVEC TNF alpha + IL4	0.5
Secondary Th2 rest	1.6	HUVEC IL-11	0.8
Secondary Tr1 rest	1.5	Lung Microvascular EC none	2.0
Primary Th1 act	0.3	Lung Microvascular EC TNFalpha + IL-1beta	0.8
Primary Th2 act	0.0	Microvascular Dermal EC none	1.9
Primary Tr1 act	0.3	Microsvascular Dermal EC TNFalpha + IL-1beta	1.7
Primary Th1 rest	4.8	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	2.1	Small airway epithelium none	0.0
Primary Tr1 rest	3.4	Small airway epithelium TNFalpha + IL-1beta	0.3
CD45RA CD4 lymphocyte act	0.6	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	1.1	Coronary artery SMC TNFalpha + IL-1beta	0.3
CD8 lymphocyte act	0.3	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	1.3	Astrocytes TNFalpha + IL-1beta	0.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.3
CD4 lymphocyte none	1.7	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.8	CCD1106 (Keratinocytes) none	0.1
LAK cells rest	2.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.4	Liver cirrhosis	0.4
LAK cells IL-2+IL-12	0.3	Lupus kidney	1.0
LAK cells IL-2+IFN gamma	0.6	NCI-H292 none	0.6
LAK cells IL-2+ IL-18	0.6	NCI-H292 IL-4	0.0
LAK cells	0.0	NCI-H292 IL-9	0.0

PMA/ionomycin			
NK Cells IL-2 rest	0.6	NCI-H292 IL-13	0.0
Two Way MLR 3 day	1.1	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	1.0	HPAEC none	0.6
Two Way MLR 7 day	0.5	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.6	Lung fibroblast none	0.0
PBMC PWM	1.1	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.2	Lung fibroblast IL-4	0.0
Ramos (B cell) none	1.7	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	1.1	Lung fibroblast IL-13	0.0
B lymphocytes PWM	1.4	Lung fibroblast IFN gamma	0.2
B lymphocytes CD40L and IL-4	4.7	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	1.4
EOL-1 dbcAMP PMA/ionomycin	0.1	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.5	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.5	IBD Colitis 2	0.2
Monocytes rest	1.2	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	4.9
Macrophages rest	0.2	Lung	3.1
Macrophages LPS	0.0	Thymus	100.0
HUVEC none	0.5	Kidney	4.6
HUVEC starved	0.3		

Panel 1.3D Summary: Ag2966 Two experiments both show that expression of the NOV24a gene, a sodium-glucose cotransporter homolog, is limited to the kidney (CTs=29). This restricted expression is in agreement with published data which has shown that secondary active transport of glucose in the kidney is mediated by sodium glucose cotransporter. (See ref. 1). Thus, expression of this gene could be used as a marker for kidney tissue. Furthermore, the protein product may be important for normal function of the kidney. Thus, therapeutic modulation of the expression or function of this protein may be useful in treating diseases that affect the kidney, including diabetes.

Bissonnette P, Noel J, Coady MJ, Lapointe JY. Functional expression of tagged human Na⁺-glucose cotransporter in *Xenopus laevis* oocytes. *J Physiol* 1999 Oct 15;520 Pt 2:359-71

1. High-affinity, secondary active transport of glucose in the intestine and kidney is mediated by an integral membrane protein named SGLT1 (sodium glucose cotransporter).
5 Though basic properties of the transporter are now defined, many questions regarding the structure- function relationship of the protein, its biosynthesis and targeting remain unanswered. In order to better address these questions, we produced a functional hSGLT1 protein (from human) containing a reporter tag. 2. Six constructs, made from three tags (myc, haemagglutinin and poly-His) inserted at both the C- and N-terminal positions, were thus tested
10 using the *Xenopus* oocyte expression system via electrophysiology and immunohistochemistry. Of these, only the hSGLT1 construct with the myc tag inserted at the N-terminal position proved to be of interest, all other constructs showing no or little transport activity. A systematic comparison of transport properties was therefore performed between the myc-tagged and the untagged hSGLT1 proteins. 3. On the basis of both steady-state (affinities
15 for substrate (glucose) and inhibitor (phlorizin) as well as expression levels) and presteady-state parameters (transient currents) we conclude that the two proteins are functionally indistinguishable, at least under these criteria. Immunological detection confirmed the appropriate targeting of the tagged protein to the plasma membrane of the oocyte with the epitope located at the extracellular side. 4. The myc-tagged hSGLT1 was also successfully
20 expressed in polarized MDCK cells. alpha-Methylglucose uptake studies on transfected cells showed an exclusively apical uptake pathway, thus indicating that the expressed protein was correctly targeted to the apical domain of the cell. 5. These comparative studies demonstrate that the myc epitope inserted at the N-terminus of hSGLT1 produces a fully functional protein while other epitopes of similar size inserted at either end of the protein inactivated the final
25 protein.

PMID: 10523405

Panel 2D Summary: Ag2966 Expression of the NOV24a gene is predominantly limited to the kidney. This result is in agreement with the expression seen in Panel 1.3D.

Panel 3D Summary: Ag2966 Results from one experiment with the NOV24a gene are
30 not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 4D Summary: Ag2966 Expression of the NOV24a gene is predominantly found in normal tissue from thymus, lung, colon and kidney. This expression profile suggests that the protein product may be involved in glucose transport in these tissues. Therefore, therapeutic

[illegible]

5

Expression of gene NOV28 was assessed using the primer-probe set Ag2891, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB, ZC, ZD and ZE.

Table ZA. Probe Name Ag2891

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccggaattaatctaccatcaa-3'	22	252	1087
Probe	TET-5'-ccccgggtaacaactgtttcagatct-3'-TAMRA	26	289	1088
Reverse	5'-gatgaaaaaqtccatgtttggt-3'	21	325	1089

Table ZB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2891, Run 224116294	Tissue Name	Rel. Exp.(%) Ag2891, Run 224116294
AD 1 Hippo	13.9	Control (Path) 3 Temporal Ctx	43.2
AD 2 Hippo	27.0	Control (Path) 4 Temporal Ctx	20.9
AD 3 Hippo	36.1	AD 1 Occipital Ctx	6.4
AD 4 Hippo	16.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	35.1	AD 3 Occipital Ctx	19.1
AD 6 Hippo	29.9	AD 4 Occipital Ctx	51.1
Control 2 Hippo	6.0	AD 5 Occipital Ctx	0.0
Control 4 Hippo	34.4	AD 6 Occipital Ctx	11.0
Control (Path) 3 Hippo	17.4	Control 1 Occipital Ctx	20.6
AD 1 Temporal Ctx	76.3	Control 2 Occipital Ctx	18.2
AD 2 Temporal Ctx	23.3	Control 3 Occipital Ctx	0.0
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	17.3
AD 4 Temporal Ctx	34.6	Control (Path) 1 Occipital Ctx	65.1
AD 5 Inf Temporal Ctx	24.5	Control (Path) 2 Occipital Ctx	13.5
AD 5 SupTemporal	58.2	Control (Path) 3	0.0

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	45.1	Control (Path) 4 Occipital Ctx	17.6
AD 6 Sup Temporal Ctx	95.9	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	5.2	Control 2 Parietal Ctx	19.2
Control 2 Temporal Ctx	0.0	Control 3 Parietal Ctx	37.9
Control 3 Temporal Ctx	12.4	Control (Path) 1 Parietal Ctx	100.0
Control 4 Temporal Ctx	21.6	Control (Path) 2 Parietal Ctx	25.9
Control (Path) 1 Temporal Ctx	65.1	Control (Path) 3 Parietal Ctx	0.0
Control (Path) 2 Temporal Ctx	55.9	Control (Path) 4 Parietal Ctx	11.2

Table ZC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2891, Run 160898914	Rel. Exp.(%) Ag2891, Run 165701349	Tissue Name	Rel. Exp.(%) Ag2891, Run 160898914	Rel. Exp.(%) Ag2891, Run 165701349
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	0.0	0.0
Pancreas	0.0	0.0	Renal ca. 786-0	72.7	17.7
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.0	0.0
Adrenal gland	0.0	0.0	Renal ca. RXF 393	98.6	100.0
Thyroid	0.0	0.0	Renal ca. ACHN	35.6	21.3
Salivary gland	0.0	0.0	Renal ca. UO-31	43.5	17.9
Pituitary gland	0.0	0.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	0.0	0.0	Liver	0.0	0.0
Brain (whole)	0.0	0.0	Liver (fetal)	0.0	0.0
Brain (amygdala)	18.4	0.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	1.9	10.2	Lung	0.0	0.0
Brain (hippocampus)	19.6	12.2	Lung (fetal)	0.0	0.0
Brain (substantia	0.0	0.0	Lung ca.	0.0	0.0

nigra)			(small cell) LX-1		
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.0
Cerebral Cortex	100.0	11.8	Lung ca. (s.cell var.) SHP-77	0.0	12.4
Spinal cord	14.9	13.5	Lung ca. (large cell)NCI-H460	0.0	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non- sm. cell) A549	0.0	26.1
glio/astro U-118- MG	0.0	0.0	Lung ca. (non- s.cell) NCI- H23	16.3	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non- s.cell) HOP-62	0.0	0.0
neuro*; met SK-N- AS	0.0	0.0	Lung ca. (non- s.cl) NCI- H522	19.8	0.0
astrocytoma SF- 539	0.0	10.5	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SNB- 75	0.0	0.0	Lung ca. (squam.) NCI- H596	0.0	0.0
glioma SNB-19	0.0	0.0	Mammary gland	0.0	37.4
glioma U251	14.4	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA- MB-231	0.0	0.0
Heart (fetal)	0.0	0.0	Breast ca.* (pl.ef) T47D	14.6	0.0
Heart	0.0	0.0	Breast ca. BT- 549	0.0	0.0
Skeletal muscle (fetal)	15.8	0.0	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.0	0.0	Ovary	58.6	0.0
Bone marrow	0.0	0.0	Ovarian ca. OVCAR-3	0.0	4.6
Thymus	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.0	0.0	Ovarian ca. OVCAR-5	0.0	0.0
Lymph node	0.0	0.0	Ovarian ca.	0.0	0.0

Tissue Name	Rel. Exp.(%) Ag2891, Run 160899401	Tissue Name	Rel. Exp.(%) Ag2891, Run 160899401
Normal Colon	0.0	Kidney Margin 8120608	18.2
CC Well to Mod Diff (ODO3866)	15.6	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	37.1
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	100.0
CC Margin (ODO3868)	0.0	Kidney Margin	6.3

		9010321	
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.0
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.0
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	20.6
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451-02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin (OD04410)	17.2	Breast Cancer Metastasis (OD04655-05)	0.0
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.0
Normal Lung 061010	24.7	Breast Cancer 9100266	0.0
Lung Met to Muscle (ODO4286)	15.7	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	0.0
Lung Margin (OD04404)	34.4	Liver Cancer 1025	8.7
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	36.6
Lung Margin (OD04565)	8.4	Liver Cancer 6004-T	8.0
Lung Cancer (OD04237-01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	8.2
Ocular Mel Met to Liver	17.0	Liver Tissue 6005-N	0.0

(ODO4310)			
Liver Margin (ODO4310)	0.0	Normal Bladder	8.3
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	45.7	Bladder Cancer A302173	10.8
Normal Kidney	41.5	Bladder Cancer (OD04718-01)	8.2
Kidney Ca, Nuclear grade 2 (OD04338)	28.3	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	47.3	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	62.9	Ovarian Cancer 064008	9.1
Kidney Margin (OD04339)	40.9	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	97.3	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	17.4	Normal Stomach	35.1
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	3.2	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	8.3	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	17.7	Stomach Margin 9060394	23.5
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	8.1
Kidney Margin (OD04450-03)	4.3	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	10.6	Gastric Cancer 064005	0.0

Table ZE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2891, Run 159632970	Tissue Name	Rel. Exp.(%) Ag2891, Run 159632970
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0

Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	3.5	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	41.8
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	1.5
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	19.5
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0

PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	7.3	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	2.2	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	4.5
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	100.0
Macrophages rest	0.0	Lung	32.3
Macrophages LPS	1.3	Thymus	40.3
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag2891 The NOV28 gene is expressed at very low levels in the CNS. No differential expression is detected in the postmortem brains of Alzheimer's patients when compared with non-demented controls. The widespread expression in the brain however suggests that this gene may be of utility in the treatment of neurological diseases.

Panel 1.3D Summary: Ag2891 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of the NOV28 gene in a lung cancer cell line and the brain (CTs=33-34). Significant expression is also seen in the testis and a cluster of lung cancer cell lines. Thus, expression of this gene could be used to differentiate these samples from other samples on this panel, and as a marker of testis tissue and lung cancer.

Panel 2D Summary: Ag2891 Expression of the NOV28 gene is limited to samples derived from kidney cancer (CTs=33-34). Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of kidney cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of kidney cancer.

Panel 3D Summary: Ag2891 Expression of the NOV28 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2891 The NOV28 transcript is expressed at low but significant levels in the colon and thymus (CTs=33-35). Thus, the transcript or the protein it encodes could be used for detection of these tissues. The protein encoded by this transcript may also play an important role in the normal homeostasis of these tissues. Therefore, therapeutics designed with the protein encoded by this transcript could be important for modulating T cell development in the thymus or maintaining or restoring normal function to these organs during inflammation due to inflammatory bowel disease in the colon.

10 NOV29a

Expression of gene NOV29a was assessed using the primer-probe set Ag2892, described in Table AAA. Results of the RTQ-PCR runs are shown in Tables AAB, AAC, AAD and AAE.

Table AAA. Probe Name Ag2892

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggcagctcgaaactttgac-3'	19	171	1090
Probe	TET-5'-cagaaatccgaagacatgctccgaag-3'-TAMRA	26	193	1091
Reverse	5'-gacaatggtgtccaggtcttgt-3'	22	240	1092

15

Table AAB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2892, Run 165721697	Tissue Name	Rel. Exp.(%) Ag2892, Run 165721697
Liver adenocarcinoma	4.6	Kidney (fetal)	4.4
Pancreas	0.7	Renal ca. 786-0	3.6
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	4.2	Renal ca. RXF 393	7.0
Thyroid	0.0	Renal ca. ACHN	2.1
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	41.5
Brain (whole)	0.6	Liver (fetal)	76.8
Brain (amygdala)	4.0	Liver ca. (hepatoblast) HepG2	8.0
Brain (cerebellum)	2.0	Lung	26.8
Brain (hippocampus)	0.0	Lung (fetal)	11.5

Brain (substantia nigra)	1.9	Lung ca. (small cell) LX-1	100.0
Brain (thalamus)	4.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	4.2	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	2.6	Lung ca. (large cell)NCI-H460	19.1
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	16.6
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	28.3
astrocytoma SW1783	2.1	Lung ca. (non-s.cell) HOP-62	10.4
neuro*; met SK-N-AS	1.3	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.7
Heart (fetal)	7.3	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	77.4	Ovarian ca. OVCAR- 3	0.0
Thymus	0.0	Ovarian ca. OVCAR- 4	2.2
Spleen	0.0	Ovarian ca. OVCAR- 5	0.0
Lymph node	0.0	Ovarian ca. OVCAR- 8	0.0
Colorectal	0.0	Ovarian ca. IGROV- 1	0.0
Stomach	5.6	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	2.7
Colon ca. SW480	32.3	Placenta	0.0
Colon ca.*	31.6	Prostate	0.0

SW620(SW480 met)			
Colon ca. HT29	1.6	Prostate ca.* (bone met)PC-3	25.0
Colon ca. HCT-116	28.9	Testis	39.2
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	1.4	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	14.2
Gastric ca.* (liver met) NCI-N87	3.4	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	4.4
Kidney	3.2	Adipose	4.1

Table AAC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2892, Run 160942674	Tissue Name	Rel. Exp.(%) Ag2892, Run 160942674
Normal Colon	1.8	Kidney Margin 8120608	0.9
CC Well to Mod Diff (ODO3866)	1.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	0.7
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	2.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	2.4	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	1.1
CC Gr.2 ascend colon (ODO3921)	1.0	Normal Thyroid	0.0
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	1.5
CC from Partial Hepatectomy (ODO4309) Mets	5.3	Thyroid Cancer A302152	1.6
Liver Margin (ODO4309)	100.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	5.6	Normal Breast	0.7
Lung Margin (OD04451-02)	4.2	Breast Cancer (OD04566)	0.0

Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin (OD04410)	0.7	Breast Cancer Metastasis (OD04655-05)	3.3
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.7	Breast Cancer 1024	0.0
Normal Lung 061010	6.7	Breast Cancer 9100266	0.0
Lung Met to Muscle (ODO4286)	3.4	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	1.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	9.2	Normal Liver	59.0
Lung Cancer (OD04404)	3.5	Liver Cancer 064003	19.2
Lung Margin (OD04404)	8.2	Liver Cancer 1025	60.7
Lung Cancer (OD04565)	1.4	Liver Cancer 1026	39.5
Lung Margin (OD04565)	12.3	Liver Cancer 6004-T	73.2
Lung Cancer (OD04237-01)	0.0	Liver Tissue 6004-N	9.8
Lung Margin (OD04237-02)	7.2	Liver Cancer 6005-T	39.5
Ocular Mel Met to Liver (ODO4310)	0.6	Liver Tissue 6005-N	7.9
Liver Margin (ODO4310)	23.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	1.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	18.8	Bladder Cancer A302173	0.0
Normal Kidney	2.0	Bladder Cancer (OD04718-01)	6.1
Kidney Ca, Nuclear grade 2 (OD04338)	1.2	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	1.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	1.8
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	0.0

Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.6
Kidney Margin (OD04340)	3.3	Normal Stomach	1.6
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.6
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	1.7
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.8	Stomach Margin 9060394	0.6
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.7
Kidney Margin (OD04450-03)	2.7	Stomach Margin 9060396	0.9
Kidney Cancer 8120607	1.4	Gastric Cancer 064005	0.0

Table AAD. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2892, Run 164629840	Tissue Name	Rel. Exp.(%) Ag2892, Run 164629840
Daoy- Medulloblastoma	5.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	2.2	JM1- pre-B-cell lymphoma	0.0
Cerebellum	3.2	Jurkat- T cell leukemia	0.0
NCI-H292-	0.0	TF-1- Erythroleukemia	0.0

Mucoepidermoid lung carcinoma			
DMS-114- Small cell lung cancer	22.8	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	14.5
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	2.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	1.7
NCI-H157- Squamous cell lung cancer (metastasis)	20.4	G401- Wilms' tumor	0.5
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	1.0
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	2.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.5
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	100.0	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	2.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.7
KM20L2- Colon cancer	0.6	PANC-1- Pancreatic epithelioid ductal carcinoma	6.5
NCI-H716- Colon cancer	0.4	T24- Bladder carcinoma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	6.6	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinoma (transitional cell)	4.4
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	2.8	HT-1080- Fibrosarcoma	0.0

NCI-SNU-5- Gastric carcinoma	63.7	MG-63- Osteosarcoma	4.3
KATO III- Gastric carcinoma	23.7	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	1.3	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	6.9
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	13.1
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	5.0
NCI-N87- Gastric carcinoma	2.1	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	8.8	CAL 27- Squamous cell carcinoma of tongue	0.0

Table AAE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2892, Run 164033139	Tissue Name	Rel. Exp.(%) Ag2892, Run 164033139
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	21.5

Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	65.1
CD45RA CD4 lymphocyte act	6.7	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	32.5
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	31.6
2ry Th1/Th2/Tr1 _anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	3.8
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	6.8
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	11.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	8.1
PBMC PHA-L	0.0	Lung fibroblast IL-4	10.4
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	4.3
B lymphocytes PWM	4.5	Lung fibroblast IFN gamma	5.8
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	10.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN	13.0

		gamma	
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	21.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	12.2
Macrophages rest	0.0	Lung	63.7
Macrophages LPS	0.0	Thymus	43.8
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2892 Highest expression of the NOV29a gene is seen in a lung cancer cell line (CT=32.7). Significant expression is also seen in a colon cancer cell line and the liver. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a diagnostic marker for the presence of colon and lung cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon and lung cancers. Results from a second run with the same probe and primer set are not included because of a potential problem in one of the sample wells.

Panel 2D Summary: Ag2892 Expression of the NOV29a gene is restricted to liver derived tissue, with highest expression in normal liver tissue (CT=32.4). Significant expression is also seen in liver cancer samples. Thus, expression of this gene could be used to differentiate liver derived samples from other samples on this panel and from other tissue samples.

Panel 3D Summary: Ag2892 Highest expression of the NOV29a gene is seen in a lung cancer cell line (CT=31.4). Significant expression is also seen in a gastric cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a diagnostic marker for the presence of gastric and lung cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of gastric and lung cancers.

Panel 4D Summary: Ag2892 Expression of the NOV29a gene is restricted to liver cirrhosis (CT=34.8). This liver specific expression is in agreement with the expression in Panels 1.3D and 2D. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel, and as a marker of liver tissue.

NOV29c

Expression of gene NOV29c was assessed using the primer-probe set Ag2893, described in Table ABA. Results of the RTQ-PCR runs are shown in Tables ABB, ABC, ABD, ABE and ABF.

Table ABA. Probe Name Ag2893

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gcccaatcctgatgactacttc-3'	22	39	1093
Probe	TET-5'-ctccaagctcggagctttgacctg-3'- TAMRA	24	73	1094
Reverse	5'-ctcagcatgtcctctgatttct-3'	22	98	1095

5

Table ABB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2893, Run 224116295	Tissue Name	Rel. Exp.(%) Ag2893, Run 224116295
AD 1 Hippo	18.6	Control (Path) 3 Temporal Ctx	15.8
AD 2 Hippo	59.0	Control (Path) 4 Temporal Ctx	43.5
AD 3 Hippo	25.3	AD 1 Occipital Ctx	41.8
AD 4 Hippo	56.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	95.9	AD 3 Occipital Ctx	27.5
AD 6 Hippo	86.5	AD 4 Occipital Ctx	58.6
Control 2 Hippo	26.1	AD 5 Occipital Ctx	47.0
Control 4 Hippo	59.5	AD 6 Occipital Ctx	31.2
Control (Path) 3 Hippo	17.3	Control 1 Occipital Ctx	20.9
AD 1 Temporal Ctx	48.6	Control 2 Occipital Ctx	63.7
AD 2 Temporal Ctx	61.1	Control 3 Occipital Ctx	62.0
AD 3 Temporal Ctx	25.2	Control 4 Occipital Ctx	28.1
AD 4 Temporal Ctx	70.7	Control (Path) 1 Occipital Ctx	82.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	24.5
AD 5 Sup Temporal Ctx	55.9	Control (Path) 3 Occipital Ctx	15.0
AD 6 Inf Temporal Ctx	86.5	Control (Path) 4 Occipital Ctx	28.3
AD 6 Sup Temporal Ctx	92.7	Control 1 Parietal Ctx	20.4

Control 1 Temporal Ctx	24.7	Control 2 Parietal Ctx	55.1
Control 2 Temporal Ctx	53.2	Control 3 Parietal Ctx	30.4
Control 3 Temporal Ctx	26.8	Control (Path) 1 Parietal Ctx	74.2
Control 3 Temporal Ctx	40.6	Control (Path) 2 Parietal Ctx	35.8
Control (Path) 1 Temporal Ctx	63.7	Control (Path) 3 Parietal Ctx	6.9
Control (Path) 2 Temporal Ctx	44.1	Control (Path) 4 Parietal Ctx	38.4

Table ABC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2893, Run 160944329	Rel. Exp.(%) Ag2893, Run 165701489	Tissue Name	Rel. Exp.(%) Ag2893, Run 160944329	Rel. Exp.(%) Ag2893, Run 165701489
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	19.5	4.3
Pancreas	2.5	4.0	Renal ca. 786-0	73.7	64.2
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.5	0.6
Adrenal gland	0.8	1.2	Renal ca. RXF 393	100.0	100.0
Thyroid	0.5	0.3	Renal ca. ACHN	28.1	18.8
Salivary gland	0.0	0.0	Renal ca. UO-31	23.5	24.3
Pituitary gland	2.0	1.8	Renal ca. TK-10	21.2	8.9
Brain (fetal)	0.9	2.0	Liver	0.0	1.3
Brain (whole)	3.6	6.3	Liver (fetal)	1.2	1.6
Brain (amygdala)	3.6	6.1	Liver ca. (hepatoblast) HepG2	10.9	8.1
Brain (cerebellum)	6.3	6.2	Lung	8.9	4.8
Brain (hippocampus)	9.3	10.4	Lung (fetal)	6.6	3.6
Brain (substantia nigra)	2.6	5.4	Lung ca. (small cell) LX-1	0.0	0.0
Brain (thalamus)	7.2	13.8	Lung ca. (small cell) NCI-H69	0.0	0.0

Cerebral Cortex	25.5	3.3	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	17.0	10.7	Lung ca. (large cell) NCI-H460	0.0	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non- sm. cell) A549	0.0	0.0
glio/astro U-118- MG	0.0	0.0	Lung ca. (non- s.cell) NCI- H23	2.3	1.1
astrocytoma SW1783	0.6	0.0	Lung ca. (non- s.cell) HOP-62	3.6	2.5
neuro*; met SK-N- AS	0.0	0.0	Lung ca. (non- s.cl) NCI- H522	0.0	0.0
astrocytoma SF- 539	0.0	0.0	Lung ca. (squam.) SW 900	0.0	2.4
astrocytoma SNB- 75	0.0	0.0	Lung ca. (squam.) NCI- H596	0.8	0.0
glioma SNB-19	0.0	0.0	Mammary gland	0.5	0.0
glioma U251	0.0	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA- MB-231	0.4	1.7
Heart (fetal)	1.4	0.0	Breast ca.* (pl.ef) T47D	4.8	0.9
Heart	0.0	0.0	Breast ca. BT- 549	0.0	0.0
Skeletal muscle (fetal)	3.3	0.0	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.4	0.0	Ovary	9.9	2.4
Bone marrow	0.0	0.3	Ovarian ca. OVCAR-3	3.2	2.2
Thymus	1.4	0.2	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	1.7	0.6	Ovarian ca. OVCAR-5	0.0	0.0
Lymph node	0.0	1.1	Ovarian ca. OVCAR-8	1.0	1.2
Colorectal	0.0	0.0	Ovarian ca. IGROV-1	0.0	0.0
Stomach	0.0	2.0	Ovarian ca.* (ascites) SK-	0.0	0.0

			OV-3		
Small intestine	0.0	2.3	Uterus	1.9	5.4
Colon ca. SW480	2.3	0.5	Placenta	0.5	0.9
Colon ca.* SW620(SW480 met)	0.0	0.6	Prostate	1.0	1.0
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC- 3	0.0	0.0
Colon ca. HCT- 116	0.0	0.0	Testis	25.7	9.3
Colon ca. CaCo-2	0.9	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	1.5	0.0	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC- 2998	0.0	0.0	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.6	Melanoma M14	0.0	0.0
Bladder	11.3	5.0	Melanoma LOX IMVI	0.0	0.0
Trachea	2.5	0.0	Melanoma* (met) SK- MEL-5	0.0	0.0
Kidney	26.2	5.3	Adipose	1.0	0.0

Table ABD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2893, Run 160966072	Tissue Name	Rel. Exp.(%) Ag2893, Run 160966072
Normal Colon	2.5	Kidney Margin 8120608	16.8
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.5	Kidney Margin 8120614	18.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	74.7
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	18.3
CC Mod Diff (ODO3920)	0.5	Normal Uterus	0.7
CC Margin (ODO3920)	0.4	Uterus Cancer 064011	10.8
CC Gr.2 ascend colon (ODO3921)	0.5	Normal Thyroid	0.0

CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	1.7
CC from Partial Hepatectomy (ODO4309) Mets	0.7	Thyroid Cancer A302152	3.4
Liver Margin (ODO4309)	0.9	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	5.0	Normal Breast	2.0
Lung Margin (OD04451- 02)	18.0	Breast Cancer (OD04566)	0.2
Normal Prostate 6546-1	0.7	Breast Cancer (OD04590-01)	0.2
Prostate Cancer (OD04410)	2.0	Breast Cancer Mets (OD04590-03)	0.8
Prostate Margin (OD04410)	8.7	Breast Cancer Metastasis (OD04655-05)	0.6
Prostate Cancer (OD04720-01)	3.8	Breast Cancer 064006	0.3
Prostate Margin (OD04720-02)	3.5	Breast Cancer 1024	0.5
Normal Lung 061010	12.9	Breast Cancer 9100266	0.5
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	1.1
Muscle Margin (ODO4286)	0.3	Breast Cancer A209073	0.2
Lung Malignant Cancer (OD03126)	8.7	Breast Margin A2090734	0.4
Lung Margin (OD03126)	21.0	Normal Liver	0.0
Lung Cancer (OD04404)	8.0	Liver Cancer 064003	0.8
Lung Margin (OD04404)	28.7	Liver Cancer 1025	0.4
Lung Cancer (OD04565)	0.7	Liver Cancer 1026	8.1
Lung Margin (OD04565)	33.0	Liver Cancer 6004-T	1.2
Lung Cancer (OD04237- 01)	1.3	Liver Tissue 6004-N	0.0
Lung Margin (OD04237- 02)	11.3	Liver Cancer 6005-T	5.4
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.3
Liver Margin (ODO4310)	0.4	Normal Bladder	17.3
Melanoma Mets to Lung (OD04321)	1.0	Bladder Cancer 1023	0.2
Lung Margin (OD04321)	32.1	Bladder Cancer A302173	1.3

Normal Kidney	42.0	Bladder Cancer (OD04718-01)	7.9
Kidney Ca, Nuclear grade 2 (OD04338)	55.5	Bladder Normal Adjacent (OD04718-03)	0.4
Kidney Margin (OD04338)	40.9	Normal Ovary	2.0
Kidney Ca Nuclear grade 1/2 (OD04339)	61.1	Ovarian Cancer 064008	10.7
Kidney Margin (OD04339)	22.8	Ovarian Cancer (OD04768-07)	0.3
Kidney Ca, Clear cell type (OD04340)	100.0	Ovary Margin (OD04768-08)	4.2
Kidney Margin (OD04340)	31.9	Normal Stomach	3.1
Kidney Ca, Nuclear grade 3 (OD04348)	3.3	Gastric Cancer 9060358	0.4
Kidney Margin (OD04348)	27.0	Stomach Margin 9060359	1.1
Kidney Cancer (OD04622-01)	30.4	Gastric Cancer 9060395	0.2
Kidney Margin (OD04622-03)	17.4	Stomach Margin 9060394	3.0
Kidney Cancer (OD04450-01)	21.5	Gastric Cancer 9060397	0.2
Kidney Margin (OD04450-03)	21.3	Stomach Margin 9060396	2.6
Kidney Cancer 8120607	70.7	Gastric Cancer 064005	0.8

Table ABE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2893, Run 165924139	Tissue Name	Rel. Exp.(%) Ag2893, Run 165924139
Daoy- Medulloblastoma	2.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0

SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.9	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	6.4	JM1- pre-B-cell lymphoma	0.0
Cerebellum	9.9	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	2.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	1.4	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	4.5	769-P- Clear cell renal carcinoma	7.7
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	100.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	73.2
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	5.3
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	1.3
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	1.6
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma	0.0

		(transitional cell)	
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinoma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	9.7	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table ABF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2893, Run 159633002	Tissue Name	Rel. Exp.(%) Ag2893, Run 159633002
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	1.2	HUVEC IL-11	0.0

Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.7	Small airway epithelium none	1.7
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	3.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.3	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	4.4
Secondary CD8 lymphocyte rest	0.2	Astrocytes TNFalpha + IL-1beta	5.8
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.8	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	1.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	12.7
LAK cells IL-2+IL-12	0.4	Lupus kidney	18.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	6.5
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	5.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	6.8
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	3.6
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	2.9
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.3	Lung fibroblast none	0.0
PBMC PWM	0.4	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0

Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.8	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.6	Dermal fibroblast CCD1070 rest	0.7
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	1.2
Monocytes LPS	0.0	Colon	0.8
Macrophages rest	0.0	Lung	34.2
Macrophages LPS	0.0	Thymus	100.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag2893 This panel does not show differential expression of the NOV29c gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for a discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag2893 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of the NOV29c gene in a renal cancer cell line (CTs=28-30). Significant expression is also seen in a cluster of renal cancer cell lines. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of renal cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of renal cancer.

This gene also is expressed at low, but significant levels in the brain. Expression of the NOV29C gene in the cerebral cortex suggests a role in CNS-specific processes. Homology to the tocopherol-associated protein (TAP) transcription factor suggests a role for NOV29C in tocopherol mediated gene transcription. Tocopherol is an essential vitamin involved in many CNS processes that may be mediated by both its antioxidant properties and ability to regulate

gene transcription via NOV29c. Genetic disruption of tocopherol processing results in tocopherol deficiency and CNS disorders such as ataxia and neurodegeneration. Agents that modulate NOV29c may thus have utility in the treatment of ataxia and neurodegenerative diseases.

5 **References:**

Yamauchi J, Iwamoto T, Kida S, Masushige S, Yamada K, Esashi T. Tocopherol-associated protein is a ligand-dependent transcriptional activator. *Biochem Biophys Res Commun* 2001 Jul 13;285(2):295-9

10 Vitamin E is a term that encompasses a group of potent, lipid-soluble, chain-breaking antioxidants. Structural analysis reveals that molecules having vitamin E activity include four isomers (alpha, beta, gamma, and delta) of both tocopherols and tocotrienols. Alpha-tocopherol has been shown to have the highest biological vitamin E activity in mammalian tissues based on fetal resorption assays, and it reverses vitamin E deficiency symptoms. Although the molecular functions fulfilled specifically by alpha-tocopherol have yet to be
15 fully described, it is unlikely that they are limited to general antioxidant functions. Here we show the functional characterization of alpha-tocopherol associated protein, TAP, which displays significant sequence similarity to the alpha-tocopherol transfer protein. Ligand competition analysis showed that recombinant TAP binds to alpha-tocopherol but not to other isomers of tocopherols. Using GFP fusion protein expression system, we observed that TAP
20 translocates from cytosol to nuclei in alpha-tocopherol-dependent fashion. Transient transfection experiment showed that TAP activates transcription of the reporter gene in alpha-tocopherol-dependent manner. These results suggest that the biological function of alpha-tocopherol is not only as an antioxidant but also as a transcriptional regulator of gene expression via association with a transcription factor TAP.

25 Yokota T, Igarashi K, Uchihara T, Jishage K, Tomita H, Inaba A, Li Y, Arita M, Suzuki H, Mizusawa H, Arai H. Delayed-onset ataxia in mice lacking alpha -tocopherol transfer protein: model for neuronal degeneration caused by chronic oxidative stress. *Proc Natl Acad Sci U S A* 2001 Dec 18;98(26):15185-90

30 alpha-Tocopherol transfer protein (alpha-TTP) maintains the concentration of serum alpha-tocopherol (vitamin E), one of the most potent fat-soluble antioxidants, by facilitating alpha-tocopherol export from the liver. Mutations of the alpha-TTP gene are linked to ataxia with isolated vitamin E deficiency (AVED). We produced a model mouse of AVED by deleting the alpha-TTP gene, which showed ataxia and retinal degeneration after 1 year of age. Because the brain alpha-TTP functions in maintaining alpha-tocopherol levels in the brain,

alpha-tocopherol was completely depleted in the alpha-TTP(-/-) mouse brain, and the neurological phenotype of alpha-TTP(-/-) mice is much more severe than that of wild-type mice when maintained on an alpha-tocopherol-deficient diet. Lipid peroxidation in alpha-TTP(-/-) mice brains showed a significant increase, especially in degenerating neurons. alpha-Tocopherol supplementation suppressed lipid peroxidation and almost completely prevented the development of neurological symptoms. This therapy almost completely corrects the abnormalities in a mouse model of human neurodegenerative disease. Moreover, alpha-TTP(-/-) mice may prove to be excellent animal models of delayed onset, slowly progressive neuronal degeneration caused by chronic oxidative stress.

Panel 2D Summary: Ag2893 Highest expression of the NOV29c gene is seen in a sample derived from a kidney cancer cell line (CT=29.5). In addition, this sample is more highly expressed in kidney cancer than in adjacent normal tissue. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of kidney cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of kidney cancer.

Panel 3D Summary: Ag2893 Expression of the NOV29c gene is detected primarily in samples derived from kidney cancer cell lines (CTs=30). Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of kidney cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of kidney cancer.

Panel 4D Summary: Ag2893 The NOV29c transcript is expressed at low but significant levels in the lung and thymus and in lupus kidney and cirrhotic liver. Thus, the transcript or the protein it encodes could be used for detection of these tissues. The expression of this gene suggests that the protein encoded by this transcript may play an important role in the normal homeostasis of the thymus and lung tissues. Therefore, therapeutics designed with the protein encoded by this transcript could be important for modulating T cell development in the thymus and for maintaining or restoring normal function to these lung during inflammation due to diseases such as asthma and emphysema. Additionally, induction of this transcript in other tissues such as the kidney and liver may be detrimental and antagonistic therapies designed with the protein encoded for by this transcript could be important in the treatment of diseases of these tissues.

NOV24b

Expression of gene NOV24b was assessed using the primer-probe set Ag1688, described in Table ACA. Results of the RTQ-PCR runs are shown in Tables ACB, ACC and ACD.

Table ACA. Probe Name Ag1688

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5' - tcagaagggaatcatgatatcg - 3'	22	577	1096
Probe	TET-5' - ccttgataaaaactccaggctcctttga- 3' - TAMRA	27	550	1097
Reverse	5' - tttggaaggtaggcatattgg - 3'	21	509	1098

Tissue Name	Rel. Exp.(%) Ag1688, Run 147249266	Tissue Name	Rel. Exp.(%) Ag1688, Run 147249266
Liver adenocarcinoma	0.0	Kidney (fetal)	9.2
Pancreas	6.7	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.2	Renal ca. A498	1.7
Adrenal gland	1.8	Renal ca. RXF 393	0.0
Thyroid	3.8	Renal ca. ACHN	0.0
Salivary gland	1.5	Renal ca. UO-31	0.0
Pituitary gland	6.1	Renal ca. TK-10	0.0
Brain (fetal)	0.5	Liver	100.0
Brain (whole)	3.6	Liver (fetal)	99.3
Brain (amygdala)	3.3	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.4	Lung	1.3
Brain (hippocampus)	6.2	Lung (fetal)	1.8
Brain (substantia nigra)	1.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	2.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	6.3	Lung ca. (s.cell var.) SHP-77	0.8
Spinal cord	3.1	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.2
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.2	Lung ca. (non-s.cl)	0.0

		NCI-H522	
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.2
astrocytoma SNB-75	0.1	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.2	Mammary gland	2.9
glioma U251	1.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.2	Breast ca.* (pl.ef) T47D	0.0
Heart	1.6	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.7	Breast ca. MDA-N	0.0
Skeletal muscle	1.2	Ovary	0.0
Bone marrow	0.5	Ovarian ca. OVCAR-3	0.2
Thymus	3.2	Ovarian ca. OVCAR-4	0.0
Spleen	1.0	Ovarian ca. OVCAR-5	0.3
Lymph node	2.9	Ovarian ca. OVCAR-8	0.0
Colorectal	0.8	Ovarian ca. IGROV-1	0.0
Stomach	3.3	Ovarian ca.* (ascites) SK-OV-3	1.0
Small intestine	6.2	Uterus	1.4
Colon ca. SW480	0.0	Placenta	0.4
Colon ca.* SW620(SW480 met)	0.0	Prostate	1.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	6.1
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	0.4
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.9
Colon ca. HCC-2998	0.2	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	4.4	Melanoma M14	0.0
Bladder	3.1	Melanoma LOX IMVI	0.0
Trachea	3.0	Melanoma* (met) SK-MEL-5	0.0

Kidney	6.8	Adipose	0.5
--------	-----	---------	-----

Table ACC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1688, Run 162646059	Tissue Name	Rel. Exp.(%) Ag1688, Run 162646059
Normal Colon	1.7	Kidney Margin 8120608	0.7
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.2	Kidney Margin 8120614	0.5
CC Gr.2 rectosigmoid (ODO3868)	0.2	Kidney Cancer 9010320	0.2
CC Margin (ODO3868)	0.1	Kidney Margin 9010321	1.0
CC Mod Diff (ODO3920)	0.1	Normal Uterus	0.2
CC Margin (ODO3920)	0.9	Uterus Cancer 064011	0.8
CC Gr.2 ascend colon (ODO3921)	0.1	Normal Thyroid	0.9
CC Margin (ODO3921)	0.1	Thyroid Cancer 064010	0.2
CC from Partial Hepatectomy (ODO4309) Mets	4.7	Thyroid Cancer A302152	0.5
Liver Margin (ODO4309)	100.0	Thyroid Margin A302153	1.0
Colon mets to lung (OD04451-01)	0.1	Normal Breast	0.3
Lung Margin (OD04451- 02)	0.1	Breast Cancer (OD04566)	0.1
Normal Prostate 6546-1	2.1	Breast Cancer (OD04590-01)	0.1
Prostate Cancer (OD04410)	0.6	Breast Cancer Mets (OD04590-03)	0.4
Prostate Margin (OD04410)	0.5	Breast Cancer Metastasis (OD04655-05)	0.9
Prostate Cancer (OD04720-01)	1.1	Breast Cancer 064006	0.6
Prostate Margin (OD04720-02)	1.6	Breast Cancer 1024	1.2
Normal Lung 061010	2.0	Breast Cancer 9100266	0.1
Lung Met to Muscle	0.0	Breast Margin	0.1

(ODO4286)		9100265	
Muscle Margin (ODO4286)	0.2	Breast Cancer A209073	0.3
Lung Malignant Cancer (OD03126)	0.1	Breast Margin A2090734	0.3
Lung Margin (OD03126)	0.5	Normal Liver	69.7
Lung Cancer (OD04404)	0.1	Liver Cancer 064003	13.7
Lung Margin (OD04404)	0.2	Liver Cancer 1025	18.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	1.2
Lung Margin (OD04565)	0.1	Liver Cancer 6004-T	22.2
Lung Cancer (OD04237-01)	0.1	Liver Tissue 6004-N	1.0
Lung Margin (OD04237-02)	0.4	Liver Cancer 6005-T	1.9
Ocular Mel Met to Liver (ODO4310)	0.1	Liver Tissue 6005-N	4.2
Liver Margin (ODO4310)	77.4	Normal Bladder	2.7
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.1	Bladder Cancer A302173	0.2
Normal Kidney	12.9	Bladder Cancer (OD04718-01)	0.1
Kidney Ca, Nuclear grade 2 (OD04338)	3.8	Bladder Normal Adjacent (OD04718-03)	0.5
Kidney Margin (OD04338)	1.6	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	2.8	Ovarian Cancer 064008	0.1
Kidney Margin (OD04339)	9.3	Ovarian Cancer (OD04768-07)	0.2
Kidney Ca, Clear cell type (OD04340)	1.4	Ovary Margin (OD04768-08)	0.1
Kidney Margin (OD04340)	4.1	Normal Stomach	0.3
Kidney Ca, Nuclear grade 3 (OD04348)	0.1	Gastric Cancer 9060358	0.1
Kidney Margin (OD04348)	3.8	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.2	Gastric Cancer 9060395	0.2
Kidney Margin (OD04622-03)	0.7	Stomach Margin 9060394	0.3
Kidney Cancer	0.2	Gastric Cancer	0.3

(OD04450-01)		9060397	
Kidney Margin (OD04450-03)	2.6	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	1.1

Table ACD. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524	Tissue Name	Rel. Exp.(%) Ag1688, Run 226587524
97457_Patient-02go_adipose	41.2	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	9.9	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	8.1	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	0.0	94712_Donor 2 AD - A_adipose	11.4
99167_Bayer Patient 1	84.7	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	2.4	94714_Donor 2 AD - C_adipose	29.1
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	19.2
97486_Patient-09sk_skeletal muscle	8.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	9.6	94730_Donor 3 AM - A_adipose	15.0
97488_Patient-09pl_placenta	0.0	94731_Donor 3 AM - B_adipose	37.9
97492_Patient-10ut_uterus	0.0	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	0.0	94733_Donor 3 AD - A_adipose	39.2
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	11.4
97496_Patient-11sk_skeletal muscle	52.9	94735_Donor 3 AD - C_adipose	34.4
97497_Patient-11ut_uterus	35.8	77138_Liver_HepG2untreated	8.4
97498_Patient-11pl_placenta	10.5	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.0	81735_Small Intestine	100.0
97501_Patient-12sk_skeletal muscle	35.4	72409_Kidney_Proximal Convoluted Tubule	9.9

97502_Patient-12ut_uterus	20.7	82685_Small intestine_Duodenum	70.2
97503_Patient-12pl_placenta	0.0	90650_Adrenal_Adrenocortical adenoma	25.5
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	10.4
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	7.2
94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0

Panel 1.3D Summary: Agl688 Expression of the NOV24b gene, a plasma kallikrein homolog, is significantly higher in liver (CTs=28) than in any other sample on this panel. Thus, expression of this gene could be used as a marker of liver tissue. Plasma kallikrein is a serine protease that, among other roles, plays a part in blood coagulation, fibrinolysis, and complement activation and has been implicated in adipose differentiation by remodelling of the fibronectin-rich ECM of preadipocytes. Therefore, an antagonist to this gene product may be beneficial in the treatment of obesity.

References:

Hoover-Plow J, Yuen L. Plasminogen binding is increased with adipocyte differentiation. Biochem.Biophys.Res.Comm. (2001) 284, 389-394

The purpose of this study was to examine the role of the plasminogen system in the development of adipose tissue. Plasminogen binding capacity was determined in differentiated and undifferentiated cells from adipose tissue of plasminogen deficient mice and 3T3 cells, a well-characterized tissue culture model. In 3T3 cells, plasminogen binding was fivefold higher in differentiated cells compared to the undifferentiated cells. Inhibition of binding by carboxyl-terminal lysine analogs was similar for the differentiated and undifferentiated cells with tranexamic acid > EACA > lysine. The binding of plasminogen was concentration-dependent and approaches saturation in the both cell types. The number of plasminogen binding sites was tenfold higher in the differentiated compared to the undifferentiated cells. In isolated mature fat cells and stromal cell cultures from mouse adipose tissue, plasminogen binding was also higher in the differentiated mature fat cells and differentiated stromal cells compared to undifferentiated stromal cells. Plasminogen binding was elevated in the differentiated cells from the Plg-/- mice compared to cells from the WT mice. These results

suggest that the plasminogen system plays an important role in adipose tissue development.
Copyright 2001 Academic Press.

PMID: 11394891

Selvarajan S, Lund LR, Takeuchi T, Craik CS, Werb Z. A plasma kallikrein-dependent
5 plasminogen cascade required for adipocyte differentiation. *Nature Cell Biol.* (2001) 3, 267-
275.

Here we show that plasma kallikrein (PKal) mediates a plasminogen (Plg) cascade in
adipocyte differentiation. Ecotin, an inhibitor of serine proteases, inhibits cell-shape change,
adipocyte-specific gene expression, and lipid accumulation during adipogenesis in culture.
10 Deficiency of Plg, but not of urokinase or tissue-type plasminogen activator, suppresses
adipogenesis during differentiation of 3T3-L1 cells and mammary-gland involution. PKal,
which is inhibited by ecotin, is required for adipose conversion, Plg activation and 3T3-L1
differentiation. Human plasma lacking PKal does not support differentiation of 3T3-L1 cells.
PKal is therefore a physiological regulator that acts in the Plg cascade during adipogenesis.
15 We propose that the Plg cascade fosters adipocyte differentiation by degradation of the
fibronectin-rich preadipocyte stromal matrix.

PMID: 11231576

Panel 2D Summary: Ag1688 The expression of the NOV24b gene appears to be
highest in a sample derived from a sample of normal liver tissue adjacent to a metastatic colon
20 cancer CT=26.2). In addition, there is substantial expression in other samples of normal liver,
and to a much lesser degree, malignant liver tissue. This liver specific expression is consistent
with the expression seen in Panel 1.3D. Thus, the expression of this gene could be used to
distinguish liver derived tissue from the other samples in the panel, and more specifically the
expression of this gene could be used to distinguish normal liver from malignant liver tissue.
25 Moreover, therapeutic modulation of this gene, through the use of small molecule drugs,
protein therapeutics or antibodies might be of benefit in the treatment of liver cancer.

Panel 5 Islet Summary: Ag1688 Expression of the NOV24b gene is limited to
pancreatic islets and small intestines. Please see Panel 1.3 for discussion of utility of this gene
in metabolic disease.

30 NOV30

Expression of gene NOV30 was assessed using the primer-probe set Ag2894,
described in Table ADA. Results of the RTQ-PCR runs are shown in Tables ADB, ADC and
ADD.

Table ADA. Probe Name Ag2894

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cctgagtcaatccaagaaactg-3'	22	94	1099
Probe	TET-5'-aggatcatcaaccaggaccgcctag-3'-TAMRA	25	116	1100
Reverse	5'-tccagtagggatctggagaagt-3'	22	149	1101

Table ADB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2894, Run 160968507	Tissue Name	Rel. Exp.(%) Ag2894, Run 160968507
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	50.3
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	8.1
Brain (whole)	0.0	Liver (fetal)	6.4
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	6.7
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	9.7
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	19.8
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	3.4	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.)	0.0

		NCI-H596	
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	3.3
Heart (fetal)	3.5	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	1.2	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	6.5
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	100.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	5.7	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	14.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	4.2

Tissue Name	Rel. Exp.(%) Ag2894, Run 160966709	Tissue Name	Rel. Exp.(%) Ag2894, Run 160966709
Normal Colon	0.0	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.0
CC Gr.2 ascend colon (ODO3921)	6.4	Normal Thyroid	0.0
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	3.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	40.3	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.0
Normal Lung 061010	9.7	Breast Cancer 9100266	0.0
Lung Met to Muscle (ODO4286)	3.9	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0

Lung Malignant Cancer (OD03126)	4.4	Breast Margin A2090734	0.0
Lung Margin (OD03126)	25.0	Normal Liver	100.0
Lung Cancer (OD04404)	12.5	Liver Cancer 064003	5.3
Lung Margin (OD04404)	15.0	Liver Cancer 1025	20.2
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	12.7
Lung Margin (OD04565)	8.1	Liver Cancer 6004-T	13.0
Lung Cancer (OD04237-01)	2.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	19.3	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	16.2	Bladder Cancer A302173	15.7
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	2.9	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	3.7	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	0.0	Normal Stomach	4.1
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	0.0	Stomach Margin 9060396	0.0

LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	13.8	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	7.6
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	23.0
Macrophages rest	0.0	Lung	13.8
Macrophages LPS	0.0	Thymus	12.8
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2894 Expression of the NOV30 gene is restricted to the testis and a renal cancer cell line(CTs=33-35). Thus, expression of this gene could be used to differentiate these samples from other samples on this panel and as a marker for testis tissue and renal cancer.

Panel 2D Summary: Ag2894 Expression of the NOV30 gene is restricted to normal liver tissue (CTs=33-35). This gene encodes a ryudocan homolog. Ryudocan is a cell-surface

heparan sulfate proteoglycan, which is involved in regulation of blood coagulation, among other biological functions. Thus, based on its expression profile, expression of this gene could be used to identify liver tissue and to differentiate between normal and malignant liver. Furthermore, this gene product may be involved in normal homeostasis of the liver. Thus, therapeutic modulation of the expression or function of this gene product may be effective in the treatment of liver disease and liver cancer.

References:

Kojima T, Inazawa J, Takamatsu J, Rosenberg RD, Saito H. Human ryudocan core protein: molecular cloning and characterization of the cDNA, and chromosomal localization of the gene. *Biochem Biophys Res Commun* 1993 Feb 15;190(3):814-22

We have isolated a series of overlapping cDNA clones encoding a 2,628 bp transcript, which potentially codes for a 198 amino acid protein with predicted molecular mass of 21,641 daltons, for the human ryudocan core protein. The deduced core proteins of the human and the rat ryudocan have high structural conservation, particularly in the NH₂ and COOH terminus regions of the putative mature core protein, including the combined transmembrane/cytoplasmic domains with conserved positions of all 4 tyrosine groups and 3 conserved glycosaminoglycan chain attachment regions, which might serve important roles for biological function of ryudocan. A major 2.7 kb transcript was detected in all tissues tested, with relatively high levels of expression observed in mRNA from lung, liver, skeletal muscle and kidney. A minor 1.9 kb transcript was also observed in some of tissues, which would be caused by alternative polyadenylation. Human ryudocan gene has been localized on the chromosome 20q12 by fluorescence in situ hybridization.

PMID: 7916598

Kojima T. Molecular biology of ryudocan, an endothelial heparan sulfate proteoglycan. *Semin Thromb Hemost* 2000;26(1):67-73

Ryudocan is a type I integral membrane heparan sulfate proteoglycan, which was originally cloned from rat microvascular endothelial cells. We have cloned the cDNA of rat ryudocan. The deduced amino acids of ryudocan has homologous transmembrane and intracellular domains with syndecan but very distinct extracellular regions. We also cloned the human ryudocan cDNA, of which the gene localizes on the chromosome 20q12. To better understand the regulation of ryudocan expression, we have determined the structural organization of the human ryudocan gene. The human ryudocan gene extends approximately 24 kb and is divided into five exons that appear conserved in syndecan family members. The 5'-flanking sequences of the human ryudocan gene contain a variety of potential binding sites

for transcription factors and are capable of functioning as a promoter. We purified human ryudocan and evaluated its interactions with several extracellular ligands. It was found that basic fibroblast growth factor (bFGF), midkine, and tissue factor pathway inhibitor exhibited significant ryudocan bindings. Heparitinase, but not chondroitin ABC lyase treatment, destroyed those ryudocan bindings; thus, the heparan sulfate chains of ryudocan appear to be responsible for those bindings. Immunohistochemical analysis revealed that ryudocan is expressed in peripheral nerve tissues, fibrous connective tissues, and placental trophoblasts. These findings suggest that ryudocan may possess multiple biologic functions, such as bFGF modulation, neurite growth promotion, and anticoagulation, via heparan sulfate-binding effectors in the cellular microenvironment.

PMID: 10805285

Panel 4D Summary: Ag2894 Expression of the NOV30 gene is restricted to a sample derived from liver cirrhosis. This gene is also expressed in normal liver in panel 2D. This expression suggests that this protein product, a ryudocan homolog, is essential to liver function. Thus, expression of this gene could be used as a marker for liver derived tissue. Furthermore, therapeutic modulation of the expression or function of this gene product may be effective in the treatment of diseases that affect the liver, including liver cirrhosis.

NOV31

Expression of gene NOV31 was assessed using the primer-probe set Ag2922, described in Table AEA. Results of the RTQ-PCR runs are shown in Table AEB.

Table AEA. Probe Name Ag2922

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cttcaggctgtgggtcatc-3'	19	407	1102
Probe	TET-5'-caagccctactgctccagtcag-3'-TAMRA	24	447	1103
Reverse	5'-cagcaacagggcttacaaca-3'	20	472	1104

Table AEB. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag2922, Run 171619741	Tissue Name	Rel. Exp.(%) Ag2922, Run 171619741
Secondary Th1 act	0.0	HUVEC IL-1beta	1.0
Secondary Th2 act	0.0	HUVEC IFN gamma	1.3
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0

Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.6	Lung Microvascular EC none	3.2
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.9
Primary Th2 act	0.4	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	1.1	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	1.3	Small airway epithelium none	0.0
Primary Tr1 rest	0.7	Small airway epithelium TNFalpha + IL-1beta	0.6
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	1.8
CD45RO CD4 lymphocyte act	1.2	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	1.6	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	2.1	Astrocytes TNFalpha + IL-1beta	0.6
Secondary CD8 lymphocyte act	1.2	KU-812 (Basophil) rest	3.4
CD4 lymphocyte none	0.4	KU-812 (Basophil) PMA/ionomycin	1.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.3	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.8	Liver cirrhosis	0.7
LAK cells IL-2+IL-12	2.0	NCI-H292 none	2.2
LAK cells IL-2+IFN gamma	1.9	NCI-H292 IL-4	4.3
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	1.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	1.8
NK Cells IL-2 rest	2.0	NCI-H292 IFN gamma	1.3
Two Way MLR 3 day	3.4	HPAEC none	0.4
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.6
Two Way MLR 7 day	0.6	Lung fibroblast none	0.6
PBMC rest	0.5	Lung fibroblast TNF alpha + IL-1 beta	0.0

PBMC PWM	0.0	Lung fibroblast IL-4	1.1
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	3.5
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	2.8
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	2.2
B lymphocytes CD40L and IL-4	2.4	Dermal fibroblast CCD1070 TNF alpha	7.3
EOL-1 dbcAMP	5.3	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	1.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	2.5	Dermal fibroblast IL-4	1.3
Dendritic cells LPS	0.9	Dermal Fibroblasts rest	0.5
Dendritic cells anti-CD40	2.3	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	1.2
Monocytes LPS	0.0	Colon	3.1
Macrophages rest	1.8	Lung	6.7
Macrophages LPS	0.9	Thymus	32.5
HUVEC none	0.8	Kidney	100.0
HUVEC starved	0.5		

CNS_neurodegeneration_v1.0 Summary: Ag2922 Expression of the NOV31 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

5 **Panel 1.3D Summary:** Ag2922 Expression of the NOV31 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 2D Summary: Ag2922 Results from one experiment with the NOV31 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

10 **Panel 4.1D Summary:** Ag2922 Expression of the NOV31 gene is restricted to normal thymus and kidney (CTs=32-33). Thus, expression of this gene could be used as a marker for kidney and thymic tissue.

Panel 4D Summary: Ag2922 Results from one experiment with the NOV31 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

NOV36a and NOV36b

Expression of gene NOV36a and variant NOV36b was assessed using the primer-probe sets Ag1136 and Ag2999, described in Tables AFA and AFB. Results of the RTQ-PCR runs are shown in Tables AFC, AFD, AFE and AFF.

Table AFA. Probe Name Ag1136

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tcaccaaagtgcagacatcaa-3'	22	455	1105
Probe	TET-5'-ttttcccttgggccctaccatg-3'- TAMRA	23	492	1106
Reverse	5'-atgtaccgacattggacatctc-3'	22	526	1107

5

Table AFB. Probe Name Ag2999

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gccctacgataacttttgtgtgt-3'	22	1488	1108
Probe	TET-5'-ctcctggccagctgattcaggtcat-3'- TAMRA	25	1520	1109
Reverse	5'-gctactgttgccaacttcattc-3'	22	1560	1110

Table AFC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag1136, Run 206992276	Tissue Name	Rel. Exp.(%) Ag1136, Run 206992276
AD 1 Hippo	14.4	Control (Path) 3 Temporal Ctx	6.2
AD 2 Hippo	44.4	Control (Path) 4 Temporal Ctx	25.7
AD 3 Hippo	7.6	AD 1 Occipital Ctx	12.2
AD 4 Hippo	6.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	91.4	AD 3 Occipital Ctx	9.0
AD 6 Hippo	48.3	AD 4 Occipital Ctx	17.7
Control 2 Hippo	35.6	AD 5 Occipital Ctx	12.5
Control,4 Hippo	14.3	AD 6 Occipital Ctx	35.4
Control (Path) 3 Hippo	6.4	Control 1 Occipital Ctx	13.6
AD 1 Temporal Ctx	16.7	Control 2 Occipital Ctx	61.1
AD 2 Temporal Ctx	31.6	Control 3 Occipital Ctx	25.5
AD 3 Temporal Ctx	15.5	Control 4 Occipital Ctx	8.4
AD 4 Temporal Ctx	21.6	Control (Path) 1 Occipital Ctx	73.7

AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	9.2
AD 5 Sup Temporal Ctx	41.8	Control (Path) 3 Occipital Ctx	4.6
AD 6 Inf Temporal Ctx	57.0	Control (Path) 4 Occipital Ctx	21.3
AD 6 Sup Temporal Ctx	37.4	Control 1 Parietal Ctx	10.6
Control 1 Temporal Ctx	16.5	Control 2 Parietal Ctx	46.0
Control 2 Temporal Ctx	45.7	Control 3 Parietal Ctx	18.7
Control 3 Temporal Ctx	20.4	Control (Path) 1 Parietal Ctx	62.9
Control 4 Temporal Ctx	12.6	Control (Path) 2 Parietal Ctx	17.9
Control (Path) 1 Temporal Ctx	51.1	Control (Path) 3 Parietal Ctx	10.1
Control (Path) 2 Temporal Ctx	31.6	Control (Path) 4 Parietal Ctx	39.5

Table AFD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1136, Run 165528214	Tissue Name	Rel. Exp.(%) Ag1136, Run 165528214
Liver adenocarcinoma	36.6	Kidney (fetal)	19.1
Pancreas	4.9	Renal ca. 786-0	30.1
Pancreatic ca. CAPAN 2	12.0	Renal ca. A498	25.9
Adrenal gland	4.7	Renal ca. RXF 393	18.6
Thyroid	4.5	Renal ca. ACHN	22.2
Salivary gland	0.6	Renal ca. UO-31	22.4
Pituitary gland	10.0	Renal ca. TK-10	31.4
Brain (fetal)	53.2	Liver	0.0
Brain (whole)	76.3	Liver (fetal)	0.9
Brain (amygdala)	43.5	Liver ca. (hepatoblast) HepG2	4.3
Brain (cerebellum)	48.3	Lung	9.7
Brain (hippocampus)	42.6	Lung (fetal)	9.7
Brain (substantia nigra)	13.9	Lung ca. (small cell) LX-1	6.7
Brain (thalamus)	37.6	Lung ca. (small cell) NCI-H69	23.3
Cerebral Cortex	28.1	Lung ca. (s.cell var.) SHP-77	49.7

Spinal cord	7.6	Lung ca. (large cell)NCI-H460	13.9
glio/astro U87-MG	12.7	Lung ca. (non-sm. cell) A549	8.4
glio/astro U-118-MG	85.9	Lung ca. (non-s.cell) NCI-H23	19.2
astrocytoma SW1783	33.9	Lung ca. (non-s.cell) HOP-62	30.4
neuro*; met SK-N-AS	72.2	Lung ca. (non-s.cl) NCI-H522	32.5
astrocytoma SF-539	51.1	Lung ca. (squam.) SW 900	47.6
astrocytoma SNB-75	39.5	Lung ca. (squam.) NCI-H596	80.7
glioma SNB-19	17.8	Mammary gland	17.2
glioma U251	100.0	Breast ca.* (pl.ef) MCF-7	38.4
glioma SF-295	13.5	Breast ca.* (pl.ef) MDA-MB-231	43.5
Heart (fetal)	12.3	Breast ca.* (pl.ef) T47D	13.8
Heart	13.5	Breast ca. BT-549	61.1
Skeletal muscle (fetal)	8.0	Breast ca. MDA-N	3.2
Skeletal muscle	0.0	Ovary	5.2
Bone marrow	1.6	Ovarian ca. OVCAR-3	19.2
Thymus	26.4	Ovarian ca. OVCAR-4	35.8
Spleen	15.1	Ovarian ca. OVCAR-5	46.0
Lymph node	23.3	Ovarian ca. OVCAR-8	54.7
Colorectal	6.3	Ovarian ca. IGROV-1	10.6
Stomach	19.8	Ovarian ca.* (ascites) SK-OV-3	24.8
Small intestine	40.1	Uterus	82.9
Colon ca. SW480	16.0	Placenta	18.0
Colon ca.* SW620(SW480 met)	4.7	Prostate	5.0
Colon ca. HT29	3.0	Prostate ca.* (bone met)PC-3	15.8
Colon ca. HCT-116	46.0	Testis	13.3
Colon ca. CaCo-2	10.1	Melanoma Hs688(A).T	3.6

Colon ca. tissue(ODO3866)	25.7	Melanoma* (met) Hs688(B).T	6.2
Colon ca. HCC-2998	52.1	Melanoma UACC-62	8.8
Gastric ca.* (liver met) NCI-N87	26.2	Melanoma M14	6.9
Bladder	9.0	Melanoma LOX IMVI	5.6
Trachea	20.6	Melanoma* (met) SK-MEL-5	0.8
Kidney	8.5	Adipose	9.9

Table AFE. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1136, Run 162599391	Tissue Name	Rel. Exp.(%) Ag1136, Run 162599391
Normal Colon	36.1	Kidney Margin 8120608	10.9
CC Well to Mod Diff (ODO3866)	29.7	Kidney Cancer 8120613	9.7
CC Margin (ODO3866)	12.2	Kidney Margin 8120614	13.8
CC Gr.2 rectosigmoid (ODO3868)	16.3	Kidney Cancer 9010320	15.3
CC Margin (ODO3868)	7.0	Kidney Margin 9010321	29.1
CC Mod Diff (ODO3920)	24.7	Normal Uterus	13.4
CC Margin (ODO3920)	10.2	Uterus Cancer 064011	42.9
CC Gr.2 ascend colon (ODO3921)	31.9	Normal Thyroid	12.8
CC Margin (ODO3921)	10.4	Thyroid Cancer 064010	20.4
CC from Partial Hepatectomy (ODO4309) Mets	9.1	Thyroid Cancer A302152	27.0
Liver Margin (ODO4309)	2.3	Thyroid Margin A302153	23.2
Colon mets to lung (OD04451-01)	22.1	Normal Breast	31.9
Lung Margin (OD04451-02)	13.2	Breast Cancer (OD04566)	37.9
Normal Prostate 6546-1	85.9	Breast Cancer (OD04590-01)	31.6
Prostate Cancer (OD04410)	34.9	Breast Cancer Mets (OD04590-03)	44.1
Prostate Margin	25.9	Breast Cancer	26.6

(OD04410)		Metastasis (OD04655-05)	
Prostate Cancer (OD04720-01)	24.1	Breast Cancer 064006	21.6
Prostate Margin (OD04720-02)	26.4	Breast Cancer 1024	21.6
Normal Lung 061010	40.3	Breast Cancer 9100266	38.7
Lung Met to Muscle (ODO4286)	100.0	Breast Margin 9100265	21.9
Muscle Margin (ODO4286)	8.2	Breast Cancer A209073	41.8
Lung Malignant Cancer (OD03126)	28.3	Breast Margin A2090734	14.4
Lung Margin (OD03126)	28.1	Normal Liver	2.2
Lung Cancer (OD04404)	51.4	Liver Cancer 064003	4.2
Lung Margin (OD04404)	5.6	Liver Cancer 1025	3.1
Lung Cancer (OD04565)	27.7	Liver Cancer 1026	8.1
Lung Margin (OD04565)	6.2	Liver Cancer 6004-T	1.3
Lung Cancer (OD04237- 01)	84.1	Liver Tissue 6004-N	3.6
Lung Margin (OD04237- 02)	10.6	Liver Cancer 6005-T	20.6
Ocular Mel Met to Liver (ODO4310)	4.5	Liver Tissue 6005-N	1.2
Liver Margin (ODO4310)	3.5	Normal Bladder	30.1
Melanoma Mets to Lung (OD04321)	19.8	Bladder Cancer 1023	16.3
Lung Margin (OD04321)	21.9	Bladder Cancer A302173	22.2
Normal Kidney	29.7	Bladder Cancer (OD04718-01)	37.1
Kidney Ca, Nuclear grade 2 (OD04338)	30.4	Bladder Normal Adjacent (OD04718- 03)	12.1
Kidney Margin (OD04338)	16.2	Normal Ovary	25.2
Kidney Ca Nuclear grade 1/2 (OD04339)	17.9	Ovarian Cancer 064008	33.2
Kidney Margin (OD04339)	18.2	Ovarian Cancer (OD04768-07)	77.9
Kidney Ca, Clear cell type (OD04340)	21.6	Ovary Margin (OD04768-08)	11.9
Kidney Margin (OD04340)	18.8	Normal Stomach	16.5

Kidney Ca, Nuclear grade 3 (OD04348)	10.4	Gastric Cancer 9060358	8.4
Kidney Margin (OD04348)	22.1	Stomach Margin 9060359	5.8
Kidney Cancer (OD04622-01)	8.4	Gastric Cancer 9060395	25.3
Kidney Margin (OD04622-03)	5.6	Stomach Margin 9060394	14.7
Kidney Cancer (OD04450-01)	42.0	Gastric Cancer 9060397	42.3
Kidney Margin (OD04450-03)	13.6	Stomach Margin 9060396	9.7
Kidney Cancer 8120607	20.2	Gastric Cancer 064005	21.9

Table AFF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1136, Run 164037277	Rel. Exp.(%) Ag2999, Run 165296355	Tissue Name	Rel. Exp.(%) Ag1136, Run 164037277	Rel. Exp.(%) Ag2999, Run 165296355
Secondary Th1 act	4.9	3.8	HUVEC IL-1beta	15.8	9.5
Secondary Th2 act	10.7	8.0	HUVEC IFN gamma	47.0	61.1
Secondary Tr1 act	12.2	10.2	HUVEC TNF alpha + IFN gamma	10.5	8.9
Secondary Th1 rest	4.9	5.8	HUVEC TNF alpha + IL4	24.1	15.1
Secondary Th2 rest	7.5	3.0	HUVEC IL-11	37.9	29.5
Secondary Tr1 rest	7.7	4.2	Lung Microvascular EC none	85.9	72.2
Primary Th1 act	5.7	7.3	Lung Microvascular EC TNFalpha + IL-1beta	37.1	33.9
Primary Th2 act	9.1	6.7	Microvascular Dermal EC none	85.9	24.7
Primary Tr1 act	12.2	9.2	Microvascular Dermal EC TNFalpha + IL-1beta	26.6	14.8
Primary Th1 rest	21.0	20.7	Bronchial epithelium TNFalpha +	21.8	15.8

			IL1beta		
Primary Th2 rest	13.6	8.5	Small airway epithelium none	3.1	0.6
Primary Tr1 rest	10.0	16.3	Small airway epithelium TNFalpha + IL-1beta	25.3	15.0
CD45RA CD4 lymphocyte act	10.7	8.2	Coronary artery SMC rest	15.3	5.7
CD45RO CD4 lymphocyte act	5.2	2.6	Coronary artery SMC TNFalpha + IL-1beta	3.1	4.6
CD8 lymphocyte act	4.5	3.2	Astrocytes rest	33.2	14.8
Secondary CD8 lymphocyte rest	2.8	3.0	Astrocytes TNFalpha + IL-1beta	18.7	15.4
Secondary CD8 lymphocyte act	6.1	2.2	KU-812 (Basophil) rest	10.2	11.0
CD4 lymphocyte none	2.4	1.6	KU-812 (Basophil) PMA/ionomycin	41.8	22.2
2ry Th1/Th2/Tr1_anti-CD95 CH11	7.5	10.0	CCD1106 (Keratinocytes) none	48.3	52.1
LAK cells rest	3.1	1.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	47.3	46.7
LAK cells IL-2	4.7	3.0	Liver cirrhosis	5.7	1.8
LAK cells IL-2+IL-12	2.5	0.8	Lupus kidney	5.0	2.8
LAK cells IL-2+IFN gamma	4.1	2.3	NCI-H292 none	22.8	11.4
LAK cells IL-2+IL-18	3.1	1.0	NCI-H292 IL-4	40.1	26.1
LAK cells PMA/ionomycin	6.3	4.3	NCI-H292 IL-9	32.1	23.2
NK Cells IL-2 rest	11.2	7.7	NCI-H292 IL-13	20.9	24.0
Two Way MLR 3 day	2.9	3.0	NCI-H292 IFN gamma	20.4	16.6
Two Way MLR 5 day	2.7	2.4	HPAEC none	67.8	66.4
Two Way MLR 7 day	2.9	0.8	HPAEC TNF alpha + IL-1 beta	16.5	17.1
PBMC rest	0.7	2.8	Lung fibroblast none	19.8	15.8

PBMC PWM	12.7	6.6	Lung fibroblast TNF alpha + IL-1 beta	5.0	6.7
PBMC PHA-L	6.7	3.2	Lung fibroblast IL-4	42.9	30.1
Ramos (B cell) none	0.8	0.0	Lung fibroblast IL-9	49.3	29.5
Ramos (B cell) ionomycin	1.0	2.2	Lung fibroblast IL-13	24.7	21.0
B lymphocytes PWM	9.7	4.3	Lung fibroblast IFN gamma	40.3	33.9
B lymphocytes CD40L and IL-4	5.2	5.0	Dermal fibroblast CCD1070 rest	75.3	52.5
EOL-1 dbcAMP	11.9	3.7	Dermal fibroblast CCD1070 TNF alpha	55.5	22.2
EOL-1 dbcAMP PMA/ionomycin	17.8	7.7	Dermal fibroblast CCD1070 IL-1 beta	18.6	18.9
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	12.3	16.0
Dendritic cells LPS	0.4	0.0	Dermal fibroblast IL-4	16.6	13.2
Dendritic cells anti- CD40	0.0	0.5	IBD Colitis 2	0.8	2.2
Monocytes rest	0.0	0.0	IBD Crohn's	1.2	1.0
Monocytes LPS	0.6	0.4	Colon	12.6	6.9
Macrophages rest	0.8	0.8	Lung	7.7	4.0
Macrophages LPS	0.0	0.4	Thymus	13.8	9.5
HUVEC none	45.4	54.0	Kidney	73.7	39.0
HUVEC starved	100.0	100.0			

CNS_neurodegeneration_v1.0 Summary: Ag1136 This panel does not show differential expression of the NOV36a gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag1136 Highest expression of the NOV36a gene, a cadherin 11 homolog, is seen in a glioma cell line U251 (CT=30.1). There is also low but significant expression in clusters of ovarian, lung, breast, kidney and colon cancer cell lines. Thus, expression of this gene could potentially be used as a diagnostic marker for these cancers. Furthermore, therapeutic inhibition of this gene product may be useful in the treatment of these cancers.

This gene also exhibits brain-preferential expression, indicating a role in CNS-specific processes. Recent research has shown that genetic deletion of cadherin-11 function acts to increase long term potentiation, a process thought to underlie learning and memory. Thus, drugs that target this gene product may have utility as memory enhancing drugs. Such drugs would have utility in treatment of CNS disorders involving memory dysfunction, including Alzheimer's disease and normal aging. In behavioral tests, Cadherin 11 deletion mutant mice show reduced fear- or anxiety-related responses. Thus, inhibitory agents targeting this gene product may also have utility as sedatives or anxiolytic agents for the treatment of CNS disorders including anxiety.

Among tissues with metabolic function, this gene has low levels of expression in pancreas, adrenal, thyroid, pituitary, adult and fetal heart, and adipose. Thus, this cadherin-like gene product may be important in the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine disease, including obesity and Types 1 and 2 diabetes. Decreased levels of cadherin have been associated with decreased insulin secretion, suggesting that increasing cadherin levels may be a potent therapeutic for Type 2 diabetes. In addition, this gene is expressed at higher levels in fetal (CT=34) vs adult skeletal muscle (CT=40) and may be useful for differentiation between the two sources of tissue. Ag2999 Results from one experiment with the CG56003-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

References:

Manabe T, Togashi H, Uchida N, Suzuki SC, Hayakawa Y, Yamamoto M, Yoda H, Miyakawa T, Takeichi M, Chisaka O. Loss of cadherin-11 adhesion receptor enhances plastic changes in hippocampal synapses and modifies behavioral responses. *Mol Cell Neurosci* 2000 Jun;15(6):534-46

Cadherins organize symmetrical junctions between the pre- and postsynaptic membranes in central synapses. One of them, cadherin-11 (cad11), is expressed in the limbic system of the brain, most strongly in the hippocampus. Immunohistochemical studies of the hippocampus showed that cad11 proteins were densely distributed in its synaptic neuropil zones; in cultured hippocampal neurons, their distribution often overlapped with that of synaptophysin, and also occasionally with that of GluR1 at spines. To assess the role of cad11 in synaptic formation and/or function, we analyzed brains of cad11-deficient mice. In these mice, long-term potentiation (LTP) in the CA1 region of the hippocampus was, unexpectedly, enhanced; and the level of LTP saturation was increased. In behavioral tests, the mutant mice showed reduced fear- or anxiety-related responses. These results suggest that the cad11-

mediated junctions may modulate synaptic efficacy, confining its dynamic changes to a limited range, or these junctions are required for normal development of synaptic organization in the hippocampus.

PMID: 10860580

- 5 Yamagata K, Nammo T, Moriwaki M, Ihara A, Iizuka K, Yang Q, Satoh T, Li M, Uenaka R, Okita K, Iwahashi H, Zhu Q, Cao Y, Imagawa A, Tochino Y, Hanafusa T, Miyagawa Ji J, Matsuzawa Y. Overexpression of Dominant-Negative Mutant Hepatocyte Nuclear Factor-1alpha in Pancreatic beta-Cells Causes Abnormal Islet Architecture With Decreased Expression of E-Cadherin, Reduced beta-cell Proliferation, and Diabetes. Diabetes. 10 2002 Jan;51(1):114-23.

One subtype of maturity-onset diabetes of the young (MODY)-3 results from mutations in the gene encoding hepatocyte nuclear factor (HNF)-1alpha. We generated transgenic mice expressing a naturally occurring dominant-negative form of human HNF-1alpha (P291fsinsC) in pancreatic beta-cells. A progressive hyperglycemia with age was seen 15 in these transgenic mice, and the mice developed diabetes with impaired glucose-stimulated insulin secretion. The pancreatic islets exhibited abnormal architecture with reduced expression of glucose transporter (GLUT2) and E-cadherin. Blockade of E-cadherin-mediated cell adhesion in pancreatic islets abolished the glucose-stimulated increases in intracellular Ca(2+) levels and insulin secretion, suggesting that loss of E-cadherin in beta-cells is 20 associated with impaired insulin secretion. There was also a reduction in beta-cell number (50%), proliferation rate (15%), and pancreatic insulin content (45%) in 2-day-old transgenic mice and a further reduction in 4-week-old animals. Our findings suggest various roles for HNF-1alpha in normal glucose metabolism, including the regulation of glucose transport, beta-cell growth, and beta-cell-to-beta-cell communication.

25 PMID: 11756330

Panel 2D Summary: Ag1136 The NOV36a gene is a good target for ovarian, gastric, breast, lung, colon, uterine and kidney cancers because it is expressed at a higher level in these cancers than the adjacent normal tissue. Therefore, expression of this gene could potentially be used as a diagnostic marker for these cancers and therapeutic inhibition may be useful in 30 treatment of these cancers.

Panel 4D Summary: Ag1136/Ag2999 Two experiments produce results that are in excellent agreement. The NOV36A transcript is expressed at low levels in hematopoietic cells and at higher levels in endothelial cells. IL-1 beta and TNFalpha treatment reduce transcript levels consistently in endothelium samples including HPAEC, HUVEC and lung

microvascular EC. Fibroblasts also express this transcript and dermal fibroblasts down regulate expression in response to IL-1 beta, gamma interferon and IL-4. This transcript encodes a putative cadherin 11 like molecule. Cadherins are adhesion molecules that regulate normal homeostasis. Loss of cadherin 11 expression can reduce the expression of factors such as VEGF-D in fibroblasts (see reference 1). Therapies designed with the protein encoded by this transcript could be important in the regulating endothelium function including leukocyte extravasation, a major component of inflammation during asthma, IBD, and psoriasis. Therapeutics designed with the protein encoded by this transcript could also be important in the treatment of osteoarthritis and osteoporosis since this protein may be important in maintaining bone density (see reference 2).

References:

Orlandini M, Oliviero S In fibroblasts Vegf-D expression is induced by cell-cell contact mediated by cadherin-11. *J Biol Chem* 2001 Mar 2;276(9):6576-81

Vascular endothelial growth factors (VEGFs) are a highly conserved family of growth factors all angiogenic in vivo with mitogenic and chemotactic activity on endothelial cells. VEGFs are expressed in fibroblasts either in hypoxia or in response to growth factors. Here we report that, differently from the other members of the family, Vegf-D is induced by cell-cell contact. By in situ hybridization we demonstrated that noninteracting fibroblasts express low levels of Vegf-D mRNA, whereas contacting cells express high levels of Vegf-D transcripts. By immunostaining we observed that the surface protein cadherin-11 is localized at the opposite sites of interacting cell surfaces. Ca(2+) deprivation from the culture medium determined the loss of cadherin-11 from the cell surfaces and down-regulation of Vegf-D mRNA. Moreover, a cadherin-11 antisense RNA construct inhibited Vegf-D expression in confluent BALB/c fibroblasts, whereas in NIH 3T3 cells, which express low levels of cadherin-11, Vegf-D induction could be obtained by overexpression of cadherin-11. This suggests that cell interaction mediated by cadherin-11 induces the expression of the angiogenic factor Vegf-D in fibroblasts.

PMID: 11108717

Kawaguchi J, Azuma Y, Hoshi K, Kii I, Takeshita S, Ohta T, Ozawa H, Takeichi M, Chisaka O, Kudo A. Targeted disruption of cadherin-11 leads to a reduction in bone density in calvaria and long bone metaphyses. *J Bone Miner Res* 2001 Jul;16(7):1265-71

The migration and adhesion of osteoblasts requires several classical cadherins. Cadherin-11, one of the classical cadherins, was expressed in mouse osteoblasts in skull bone and femur, revealed by immunohistochemistry. To elucidate the function of cadherin-11 in

osteoblastogenesis, cadherin-11 null mutant mice were investigated. Although apparently normal at birth, Alizarin red staining of null mutant mice showed a reduced calcified area at the frontal suture that caused a round-shaped calvaria with increasing animal age to 3 months. Consequently, there was a reduction in bone density at the femoral metaphyses and the diploe
5 of calvaria in null mutant mice. In the in vitro culture of newborn calvarial cells, the calcified area of mutant cells was smaller than those derived from wild-type littermates. These results show that absence of cadherin-11 leads to reduced bone density in some parts of skeletons including calvaria and long bone metaphyses, and thus suggest that cadherin-11 plays roles in the regulation of osteoblast differentiation and in the mineralization of the osteoid matrix.

PMID: 11450702

NOV37

Expression of gene NOV37 was assessed using the primer-probe sets Ag047, Ag2679, Ag2728, Ag332, Ag47b, Ag712, Ag2732 and Ag2975, described in Tables AGA, AGB, AGC, AGD, AGE, AGF, AGG and AGH. Results of the RTQ-PCR runs are shown in Tables AGI,
15 AGJ, AGK, AGL, AGM, AGN, AGO, AGP, AGQ and AGR.

Table AGA. Probe Name Ag047

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ccaatgacctggccacca-3'	18	1064	1111
Probe	TET-5'-ccagagtcggttcagcttcaggacagc-3'-TAMRA	27	1084	1112
Reverse	5'-gtggcacggttgctgttttagc-3'	20	1116	1113

Table AGB. Probe Name Ag2679

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ttgacctcaggaacgggtttac-3'	21	1213	1114
Probe	TET-5'-ctgctgccaggaatactttctccag-3'-TAMRA	26	1249	1115
Reverse	5'-agtatttgaggggcttcttcag-3'	22	1288	1116

Table AGC. Probe Name Ag2728

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggagcttggtctcatgacctta-3'	21	5178	1117
Probe	TET-5'-actgggctcctggccaccaagag-3'-TAMRA	23	5209	1118
Reverse	5'-agtctccatcctgtttcatc-3'	21	5233	1119

Table AGD. Probe Name Ag332

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gctgccctgacttgtgcaa-3'	19	2594	1120
Probe	TET-5'-tctgacccagtgtgcatctcccgtt-3'- TAMRA	25	2617	1121
Reverse	5'-ccggctctggcagacacact-3'	19	2651	1122

Table AGE. Probe Name Ag47b

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaacgccggagcatacaga-3'	19	1774	1123
Probe	TET-5'-ccaggtactgcacaaacacggcttcat- 3'-TAMRA	27	1805	1124
Reverse	5'-gatgccacaggccaca-3'	17	1833	1125

5

Table AGF. Probe Name Ag712

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgcaagggtacgaggatat-3'	20	8369	1126
Probe	TET-5'-cttcccgtggagcaataccagag-3'- TAMRA	24	8395	1127
Reverse	5'-tggatgttgctgctactgtct-3'	21	8424	1128

Table AGG. Probe Name Ag2732

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggagcttgggtctcatgaccta-3'	21	5178	1129
Probe	TET-5'-actgggctcctggccaccaagag-3'- TAMRA	23	5209	1130
Reverse	5'-agtcgtccatcctgtttcatc-3'	21	5233	1131

Table AGH. Probe Name Ag2975

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggagcttgggtctcatgaccta-3'	21	5178	1132
Probe	TET-5'-actgggctcctggccaccaagag-3'- TAMRA	23	5209	1133
Reverse	5'-agtcgtccatcctgtttcatc-3'	21	5233	1134

10

Table AGI. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag047, Run	Rel. Exp.(%) Ag2679, Run	Rel. Exp.(%) Ag2728, Run	Tissue Name	Rel. Exp.(%) Ag047, Run	Rel. Exp.(%) Ag2679, Run	Rel. Exp.(%) Ag2728, Run

	206228021	209751329	206483376		206228021	209751329	206483376
AD 1 Hippo	10.3	6.6	8.4	Control (Path) 3 Temporal Ctx	4.1	3.2	2.3
AD 2 Hippo	18.0	20.9	17.3	Control (Path) 4 Temporal Ctx	13.8	33.9	29.7
AD 3 Hippo	6.6	4.6	6.4	AD 1 Occipital Ctx	18.8	16.8	17.1
AD 4 Hippo	5.6	4.4	7.7	AD 2 Occipital Ctx (Missing)	0.0	0.0	0.0
AD 5 Hippo	100.0	100.0	95.9	AD 3 Occipital Ctx	5.7	3.4	3.1
AD 6 Hippo	37.6	33.0	31.4	AD 4 Occipital Ctx	12.9	15.1	19.5
Control 2 Hippo	22.1	17.3	43.2	AD 5 Occipital Ctx	25.2	33.4	51.4
Control 4 Hippo	7.7	6.4	5.3	AD 6 Occipital Ctx	3.2	18.9	19.8
Control (Path) 3 Hippo	1.5	3.4	3.7	Control 1 Occipital Ctx	3.4	3.5	4.4
AD 1 Temporal Ctx	8.1	6.7	8.6	Control 2 Occipital Ctx	83.5	82.9	100.0
AD 2 Temporal Ctx	21.2	40.3	21.2	Control 3 Occipital Ctx	25.3	20.2	17.3
AD 3 Temporal Ctx	3.3	4.2	4.2	Control 4 Occipital Ctx	5.4	4.2	4.4
AD 4 Temporal Ctx	18.7	13.1	20.0	Control (Path) 1 Occipital Ctx	66.4	74.7	75.8
AD 5 Inf Temporal Ctx	27.0	57.4	42.9	Control (Path) 2 Occipital Ctx	15.2	11.3	13.0

Salivary gland	0.0	0.2	0.0	0.0	0.0
Pituitary gland	0.0	0.1	2.1	0.0	0.0
Brain (fetal)	0.0	15.0	3.4	11.0	21.8
Brain (whole)	95.9	67.8	3.2	4.5	32.5
Brain (amygdala)	0.0	8.8	3.8	8.7	12.2
Brain (cerebellum)	0.0	22.2	1.4	0.1	14.3
Brain (hippocampus)	0.0	24.7	3.4	8.7	15.6
Brain (substantia nigra)	3.4	3.6	1.3	2.3	3.8
Brain (thalamus)	3.5	4.7	3.3	5.0	9.1
Brain (hypothalamus)	0.0	0.7	0.2	0.0	0.0
Spinal cord	0.7	1.5	0.9	1.4	1.1
glio/astro U87-MG	0.6	2.6	1.6	1.7	3.0
glio/astro U-118-MG	0.0	0.6	0.2	1.1	0.1
astrocytoma SW1783	33.7	29.5	9.0	42.9	36.1
neuro*; met SK-N-AS	0.0	0.0	0.0	0.0	0.0
astrocytoma SF-539	31.0	36.9	10.3	68.3	48.3
astrocytoma SNB-75	33.7	32.8	7.1	23.0	50.3
glioma SNB-19	100.0	100.0	30.1	100.0	100.0
glioma U251	49.0	44.1	16.5	57.8	41.5
glioma SF-295	6.0	8.1	2.1	19.8	8.7
Heart	61.1	26.8	36.6	70.7	39.8
Skeletal muscle	0.0	0.1	0.2	0.0	0.0
Bone marrow	0.0	0.1	0.0	0.0	0.0
Thymus	18.7	18.4	1.1	2.7	17.2
Spleen	0.0	0.0	0.0	0.0	0.0
Lymph node	0.0	0.2	0.0	0.0	0.0
Colon (ascending)	0.5	0.7	0.7	0.9	4.5
Stomach	0.1	1.1	0.2	0.2	0.2
Small intestine	0.0	0.1	0.0	0.0	0.0
Colon ca. SW480	0.6	1.0	1.6	7.0	1.4
Colon ca.* SW620 (SW480 met)	0.0	0.0	0.0	0.0	0.0
Colon ca. HT29	0.0	0.0	0.0	0.0	0.0

Colon ca. HCT-116	0.0	0.0	0.0	0.0	0.0
Colon ca. CaCo-2	0.0	0.1	0.0	0.0	0.0
Colon ca. HCT-15	0.0	0.1	0.2	0.0	0.0
Colon ca. HCC-2998	0.0	0.0	0.0	0.0	0.0
Gastric ca. * (liver met) NCI-N87	0.0	0.4	0.2	0.0	0.0
Bladder	0.3	1.0	0.2	0.0	0.1
Trachea	0.0	0.4	0.2	0.0	0.0
Kidney	0.2	0.9	0.7	1.2	1.0
Kidney (fetal)	1.3	3.9	0.8	3.2	2.7
Renal ca. 786-0	10.6	11.7	3.4	7.9	12.2
Renal ca. A498	0.0	0.1	0.0	0.0	0.0
Renal ca. RXF 393	17.9	14.0	4.1	16.6	15.2
Renal ca. ACHN	0.0	0.0	0.0	0.0	0.0
Renal ca. UO-31	0.0	0.2	0.1	0.0	0.2
Renal ca. TK-10	0.0	0.0	0.0	0.0	0.0
Liver	0.0	3.2	0.8	0.9	1.9
Liver (fetal)	0.0	0.1	0.0	0.0	0.0
Liver ca. (hepatoblast) HepG2	0.0	0.0	0.0	0.0	0.0
Lung	0.0	0.2	0.0	0.0	0.0
Lung (fetal)	0.0	0.5	0.0	0.0	0.0
Lung ca. (small cell) LX-1	0.0	0.0	0.0	0.0	0.0
Lung ca. (small cell) NCI-H69	2.0	3.3	1.4	5.4	2.9
Lung ca. (s.cell var.) SHP-77	0.0	0.0	0.0	0.0	0.0
Lung ca. (large cell) NCI-H460	0.0	0.0	1.3	0.7	0.3
Lung ca. (non-sm. cell) A549	0.0	0.0	0.1	0.0	0.0
Lung ca. (non-s.cell) NCI-H23	0.0	0.0	0.0	0.0	0.0
Lung ca. (non-s.cell) HOP-62	5.2	4.7	3.9	11.3	6.7
Lung ca. (non-s.cl) NCI-H522	1.7	3.1	1.5	2.8	4.8
Lung ca. (squam.) SW 900	0.0	0.0	0.0	0.0	0.0
Lung ca. (squam.)	1.3	2.5	2.0	4.2	3.0

NCI-H596					
Mammary gland	11.9	9.1	4.6	7.0	10.6
Breast ca.* (pl.ef) MCF-7	0.0	0.0	0.0	0.0	0.0
Breast ca.* (pl.ef) MDA-MB-231	0.0	0.1	0.0	0.0	0.0
Breast ca.* (pl. ef) T47D	11.8	7.5	7.9	31.6	31.4
Breast ca. BT-549	0.0	0.0	4.5	13.8	28.3
Breast ca. MDA- N	0.0	0.1	0.0	0.0	0.0
Ovary	0.1	0.6	0.4	0.1	0.2
Ovarian ca. OVCA-3	0.0	0.0	0.0	0.0	0.0
Ovarian ca. OVCA-4	0.0	0.1	1.3	4.4	0.0
Ovarian ca. OVCA-5	73.2	38.2	14.4	57.0	36.1
Ovarian ca. OVCA-8	0.0	0.7	1.0	3.4	0.2
Ovarian ca. IGROV-1	0.0	0.0	0.8	2.6	0.0
Ovarian ca. (ascites) SK-OV-3	0.0	0.0	0.0	0.0	0.0
Uterus	0.4	1.2	0.6	1.7	2.9
Placenta	0.0	0.1	0.0	0.0	0.0
Prostate	0.9	1.9	0.7	1.9	1.7
Prostate ca.* (bone met) PC-3	0.0	0.0	0.1	0.0	0.0
Testis	25.7	22.2	0.7	0.7	26.4
Melanoma Hs688(A).T	23.8	21.6	5.0	20.0	28.9
Melanoma* (met) Hs688(B).T	4.6	6.5	3.4	9.6	9.2
Melanoma UACC-62	0.0	0.0	0.0	0.0	0.0
Melanoma M14	0.0	0.1	0.0	0.0	0.0
Melanoma LOX IMVI	3.7	3.4	0.6	0.2	2.7
Melanoma* (met) SK-MEL-5	0.0	0.0	0.0	0.0	0.0
Melanoma SK- MEL-28	0.0	0.0	3.6	17.4	0.0

Table AGK. Panel 1.1

Tissue Name	Rel. Exp.(%) Ag047, Run 109663520	Tissue Name	Rel. Exp.(%) Ag047, Run 109663520
Adrenal gland	0.5	Renal ca. UO-31	0.1
Bladder	1.2	Renal ca. RXF 393	8.7
Brain (amygdala)	2.8	Liver	1.3
Brain (cerebellum)	4.1	Liver (fetal)	0.0
Brain (hippocampus)	9.6	Liver ca. (hepatoblast) HepG2	0.0
Brain (substantia nigra)	15.0	Lung	0.1
Brain (thalamus)	4.7	Lung (fetal)	0.4
Cerebral Cortex	48.3	Lung ca. (non-s.cell) HOP-62	26.6
Brain (fetal)	21.2	Lung ca. (large cell)NCI-H460	1.2
Brain (whole)	9.0	Lung ca. (non-s.cell) NCI-H23	0.0
glio/astro U-118-MG	0.7	Lung ca. (non-s.cl) NCI-H522	5.8
astrocytoma SF-539	38.4	Lung ca. (non-sm. cell) A549	0.0
astrocytoma SNB-75	16.4	Lung ca. (s.cell var.) SHP-77	0.0
astrocytoma SW1783	19.1	Lung ca. (small cell) LX-1	0.0
glioma U251	55.5	Lung ca. (small cell) NCI-H69	5.3
glioma SF-295	8.7	Lung ca. (squam.) SW 900	0.0
glioma SNB-19	100.0	Lung ca. (squam.) NCI-H596	5.0
glio/astro U87-MG	3.3	Lymph node	0.3
neuro*; met SK-N-AS	0.0	Spleen	0.0
Mammary gland	2.0	Thymus	0.6
Breast ca. BT-549	6.4	Ovary	0.6
Breast ca. MDA-N	0.0	Ovarian ca. IGROV- 1	0.0
Breast ca.* (pl.ef) T47D	5.7	Ovarian ca. OVCAR- 3	0.0
Breast ca.* (pl.ef) MCF-7	0.0	Ovarian ca. OVCAR- 4	0.1
Breast ca.* (pl.ef) MDA-MB-231	0.0	Ovarian ca. OVCAR- 5	49.0
Small intestine	0.0	Ovarian ca. OVCAR-	0.5

		8	
Colorectal	0.1	Ovarian ca.* (ascites) SK-OV-3	0.1
Colon ca. HT29	0.0	Pancreas	0.6
Colon ca. CaCo-2	0.0	Pancreatic ca. CAPAN 2	0.0
Colon ca. HCT-15	0.1	Pituitary gland	0.5
Colon ca. HCT-116	0.0	Placenta	0.0
Colon ca. HCC-2998	0.0	Prostate	0.8
Colon ca. SW480	1.2	Prostate ca.* (bone met) PC-3	0.1
Colon ca.* SW620 (SW480 met)	0.0	Salivary gland	0.2
Stomach	0.2	Trachea	0.5
Gastric ca. (liver met) NCI-N87	0.1	Spinal cord	1.4
Heart	80.7	Testis	0.5
Skeletal muscle (Fetal)	1.6	Thyroid	0.3
Skeletal muscle	0.4	Uterus	0.0
Endothelial cells	0.0	Melanoma M14	0.0
Heart (Fetal)	14.2	Melanoma LOX IMVI	0.7
Kidney	2.4	Melanoma UACC-62	0.0
Kidney (fetal)	1.5	Melanoma SK-MEL- 28	0.0
Renal ca. 786-0	3.6	Melanoma* (met) SK-MEL-5	0.0
Renal ca. A498	0.0	Melanoma Hs688(A).T	16.5
Renal ca. ACHN	0.0	Melanoma* (met) Hs688(B).T	5.6
Renal ca. TK-10	0.0		

Table AGL. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag712, Run 114986148	Rel. Exp.(%) Ag712, Run 119452123	Tissue Name	Rel. Exp.(%) Ag712, Run 114986148	Rel. Exp.(%) Ag712, Run 119452123
Endothelial cells	0.0	0.0	Renal ca. 786- 0	4.0	3.0
Heart (Fetal)	4.0	2.1	Renal ca. A498	0.1	0.1
Pancreas	2.1	1.1	Renal ca. RXF 393	4.9	3.5
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. ACHN	0.6	0.3

Adrenal Gland	2.8	2.2	Renal ca. UO-31	0.3	0.1
Thyroid	2.1	1.5	Renal ca. TK-10	0.2	0.1
Salivary gland	1.9	1.1	Liver	4.6	2.9
Pituitary gland	4.6	3.7	Liver (fetal)	0.4	0.3
Brain (fetal)	22.2	29.3	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (whole)	33.2	26.1	Lung	0.8	0.6
Brain (amygdala)	18.6	12.9	Lung (fetal)	1.3	1.2
Brain (cerebellum)	2.3	1.8	Lung ca. (small cell) LX-1	0.0	0.0
Brain (hippocampus)	27.5	18.0	Lung ca. (small cell) NCI-H69	4.4	3.7
Brain (thalamus)	10.4	7.7	Lung ca. (s.cell var.) SHP-77	0.1	0.0
Cerebral Cortex	25.3	22.2	Lung ca. (large cell) NCI-H460	1.0	0.6
Spinal cord	3.8	3.2	Lung ca. (non-sm. cell) A549	0.3	0.1
glio/astro U87-MG	2.5	2.1	Lung ca. (non-s.cell) NCI-H23	0.1	0.1
glio/astro U-118-MG	0.8	0.4	Lung ca. (non-s.cell) HOP-62	7.2	5.8
astrocytoma SW1783	19.9	9.2	Lung ca. (non-s.cl) NCI-H522	5.9	4.0
neuro*; met SK-N-AS	0.1	0.0	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SF-539	36.1	19.8	Lung ca. (squam.) NCI-H596	4.6	3.7
astrocytoma SNB-75	10.7	3.6	Mammary gland	7.6	4.5
glioma SNB-19	71.2	51.4	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma U251	38.7	19.9	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
glioma SF-295	4.6	4.1	Breast ca.* (pl.ef) T47D	10.2	7.8
Heart	100.0	100.0	Breast ca. BT-549	8.4	5.5

Skeletal Muscle	1.4	1.1	Breast ca. MDA-N	0.5	0.1
Bone marrow	0.3	0.0	Ovary	0.5	0.3
Thymus	2.1	1.0	Ovarian ca. OVCAR-3	0.4	0.1
Spleen	0.0	0.0	Ovarian ca. OVCAR-4	1.8	1.4
Lymph node	0.8	0.3	Ovarian ca. OVCAR-5	39.0	30.1
Colorectal Tissue	0.0	0.0	Ovarian ca. OVCAR-8	0.6	0.3
Stomach	1.2	0.7	Ovarian ca. IGROV-1	8.8	8.3
Small intestine	0.8	0.4	Ovarian ca. (ascites) SK- OV-3	0.1	0.0
Colon ca. SW480	2.5	1.7	Uterus	0.3	0.2
Colon ca.* SW620 (SW480 met)	0.0	0.0	Placenta	0.3	0.0
Colon ca. HT29	0.0	0.0	Prostate	5.1	3.2
Colon ca. HCT- 116	0.0	0.0	Prostate ca.* (bone met) PC- 3	0.1	0.0
Colon ca. CaCo- 2	0.3	0.2	Testis	9.7	6.8
Colon ca. Tissue (ODO3866)	0.1	0.0	Melanoma Hs688(A).T	12.0	10.4
Colon ca. HCC- 2998	0.0	0.0	Melanoma* (met) Hs688(B).T	5.0	4.3
Gastric ca.* (liver met) NCI- N87	0.5	0.2	Melanoma UACC-62	0.0	0.0
Bladder	0.9	0.6	Melanoma M14	0.0	0.0
Trachea	1.2	0.9	Melanoma LOX IMVI	0.7	0.2
Kidney	4.0	2.7	Melanoma* (met) SK- MEL-5	1.0	0.5
Kidney (fetal)	6.0	0.0			

Table AGM. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2679, Run 158633802	Rel. Exp.(%) Ag2728, Run 158560797	Rel. Exp.(%) Ag2732, Run 162400878	Tissue Name	Rel. Exp.(%) Ag2679, Run 158633802	Rel. Exp.(%) Ag2728, Run 158560797	Rel. Exp.(%) Ag2732, Run 162400878
Liver adenocarcinoma	3.8	2.0	2.9	Kidney (fetal)	0.2	1.6	1.2
Pancreas	0.0	0.0	0.0	Renal ca. 786-0	9.8	11.5	3.3
Pancreatic ca. CAPAN 2	0.0	0.0	0.0	Renal ca. A498	54.7	40.6	15.1
Adrenal gland	0.2	0.5	0.2	Renal ca. RXF 393	11.9	7.3	8.2
Thyroid	0.6	1.3	0.2	Renal ca. ACHN	0.0	0.0	0.0
Salivary gland	0.1	0.3	0.1	Renal ca. UO-31	0.8	0.3	0.0
Pituitary gland	0.4	0.3	0.0	Renal ca. TK-10	0.0	0.0	0.0
Brain (fetal)	12.4	15.2	2.5	Liver	0.7	0.4	0.4
Brain (whole)	7.2	15.5	6.1	Liver (fetal)	0.2	0.4	0.0
Brain (amygdala)	11.3	17.4	7.0	Liver ca. (hepatoblast) HepG2	0.0	0.0	0.0
Brain (cerebellum)	1.3	1.0	0.4	Lung	0.1	0.8	0.1
Brain (hippocampus)	60.7	80.1	16.8	Lung (fetal)	0.4	0.6	0.4
Brain (substantia nigra)	1.7	1.7	0.5	Lung ca. (small cell) LX-1	0.0	0.0	0.0
Brain (thalamus)	3.7	10.4	4.5	Lung ca. (small cell) NCI-H69	3.9	14.0	3.7
Cerebral Cortex	100.0	100.0	93.3	Lung ca. (s.cell var.) SHP-77	0.0	0.2	0.0
Spinal cord	1.8	1.8	1.3	Lung ca. (large cell) NCI-H460	0.2	0.1	0.5

glio/astro U87-MG	3.0	4.7	3.0	Lung ca. (non-sm. cell) A549	0.0	0.2	0.0
glio/astro U-118-MG	5.3	5.4	0.8	Lung ca. (non-s.cell) NCI-H23	0.0	0.0	0.0
astrocytom a SW1783	99.3	87.7	100.0	Lung ca. (non-s.cell) HOP-62	6.4	4.3	3.6
neuro*; met SK-N- AS	0.1	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	1.3	1.5	0.9
astrocytom a SF-539	80.7	75.3	42.0	Lung ca. (squam.) SW 900	0.0	0.0	0.0
astrocytom a SNB-75	77.4	64.6	18.2	Lung ca. (squam.) NCI-H596	1.0	2.5	2.6
glioma SNB-19	94.0	75.3	63.7	Mammary gland	1.9	4.3	0.6
glioma U251	68.8	62.4	29.1	Breast ca.* (pl.ef) MCF-7	0.0	0.0	0.0
glioma SF- 295	8.1	6.9	4.0	Breast ca.* (pl.ef) MDA-MB- 231	0.8	0.5	0.1
Heart (fetal)	34.6	10.6	11.9	Breast ca.* (pl.ef) T47D	15.9	7.6	4.9
Heart	12.9	12.1	20.2	Breast ca. BT-549	58.2	36.6	4.0
Skeletal muscle (fetal)	11.0	7.0	4.1	Breast ca. MDA-N	0.2	0.0	0.0
Skeletal muscle	0.0	0.5	0.0	Ovary	1.8	1.5	1.4
Bone marrow	0.4	0.2	0.1	Ovarian ca. OVCAR-3	0.0	0.0	0.0
Thymus	0.9	1.2	3.8	Ovarian ca. OVCAR-4	0.1	0.5	0.2
Spleen	0.0	0.0	0.0	Ovarian ca. OVCAR-5	41.8	33.9	14.3
Lymph node	0.1	0.6	0.0	Ovarian ca. OVCAR-8	1.0	0.6	0.5
Colorectal	1.3	0.5	0.9	Ovarian ca. IGROV-1	0.0	4.5	1.8
Stomach	0.4	0.5	0.1	Ovarian	0.1	0.1	0.0

				ca.* (ascites) SK-OV-3			
Small intestine	0.0	0.6	0.2	Uterus	0.0	0.0	0.0
Colon ca. SW480	11.2	5.4	1.1	Placenta	0.1	0.0	0.0
Colon ca.* SW620(S W480 met)	0.0	0.0	0.0	Prostate	1.2	0.5	0.6
Colon ca. HT29	0.0	0.0	0.0	Prostate ca.* (bone met)PC-3	0.5	0.0	0.0
Colon ca. HCT-116	0.0	0.0	0.0	Testis	1.8	1.4	0.9
Colon ca. CaCo-2	0.4	0.1	0.1	Melanoma Hs688(A). T	42.3	22.4	17.1
Colon ca. tissue(OD O3866)	0.4	0.4	0.3	Melanoma * (met) Hs688(B). T	1.7	1.6	1.0
Colon ca. HCC-2998	0.0	0.0	0.0	Melanoma UACC-62	0.0	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.2	0.4	0.1	Melanoma M14	0.0	0.0	0.0
Bladder	0.5	0.3	0.3	Melanoma LOX IMVI	2.4	1.1	0.4
Trachea	1.3	1.5	1.0	Melanoma * (met) SK-MEL-5	0.0	0.0	0.0
Kidney	0.7	0.3	0.2	Adipose	0.2	0.7	0.0

Table AGN. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag2975, Run 173763053	Tissue Name	Rel. Exp.(%) Ag2975, Run 173763053
Normal Colon	2.5	Kidney Margin (OD04348)	29.7
Colon cancer (OD06064)	4.7	Kidney malignant cancer (OD06204B)	0.0
Colon Margin (OD06064)	4.3	Kidney normal adjacent tissue (OD06204E)	0.0
Colon cancer	2.2	Kidney Cancer	0.0

(OD06159)		(OD04450-01)	
Colon Margin (OD06159)	7.3	Kidney Margin (OD04450-03)	13.6
Colon cancer (OD06297-04)	0.0	Kidney Cancer 8120613	2.3
Colon Margin (OD06297-015)	0.0	Kidney Margin 8120614	11.5
CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer 9010320	23.8
CC Margin (ODO3921)	0.0	Kidney Margin 9010321	20.0
Colon cancer metastasis (OD06104)	0.0	Kidney Cancer 8120607	2.0
Lung Margin (OD06104)	0.0	Kidney Margin 8120608	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Uterus	4.4
Lung Margin (OD04451-02)	2.0	Uterine Cancer 064011	0.0
Normal Prostate	7.5	Normal Thyroid	4.4
Prostate Cancer (OD04410)	2.3	Thyroid Cancer 064010	0.0
Prostate Margin (OD04410)	4.6	Thyroid Cancer A302152	2.2
Normal Ovary	2.2	Thyroid Margin A302153	3.6
Ovarian cancer (OD06283-03)	0.0	Normal Breast	55.1
Ovarian Margin (OD06283-07)	0.0	Breast Cancer (OD04566)	3.9
Ovarian Cancer 064008	13.4	Breast Cancer 1024	68.3
Ovarian cancer (OD06145)	18.8	Breast Cancer (OD04590-01)	5.3
Ovarian Margin (OD06145)	8.1	Breast Cancer Mets (OD04590-03)	0.0
Ovarian cancer (OD06455-03)	5.0	Breast Cancer Metastasis (OD04655-05)	13.1
Ovarian Margin (OD06455-07)	2.3	Breast Cancer 064006	12.5
Normal Lung	2.3	Breast Cancer 9100266	34.6
Invasive poor diff. lung adeno (ODO4945-01)	0.0	Breast Margin 9100265	17.0
Lung Margin (ODO4945-03)	0.0	Breast Cancer A209073	16.7
Lung Malignant Cancer	1.4	Breast Margin	71.7

(OD03126)		A2090734	
Lung Margin (OD03126)	0.0	Breast cancer (OD06083)	20.7
Lung Cancer (OD05014A)	0.0	Breast cancer node metastasis (OD06083)	6.3
Lung Margin (OD05014B)	2.1	Normal Liver	0.0
Lung cancer (OD06081)	100.0	Liver Cancer 1026	8.1
Lung Margin (OD06081)	4.1	Liver Cancer 1025	29.5
Lung Cancer (OD04237-01)	0.0	Liver Cancer 6004-T	8.4
Lung Margin (OD04237-02)	0.0	Liver Tissue 6004-N	4.7
Ocular Melanoma Metastasis	6.7	Liver Cancer 6005-T	23.0
Ocular Melanoma Margin (Liver)	2.8	Liver Tissue 6005-N	32.8
Melanoma Metastasis	2.2	Liver Cancer 064003	12.8
Melanoma Margin (Lung)	3.7	Normal Bladder	0.0
Normal Kidney	9.0	Bladder Cancer 1023	6.3
Kidney Ca, Nuclear grade 2 (OD04338)	19.5	Bladder Cancer A302173	16.7
Kidney Margin (OD04338)	3.1	Normal Stomach	15.3
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04339)	9.3	Stomach Margin 9060396	3.8
Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer 9060395	6.5
Kidney Margin (OD04340)	4.6	Stomach Margin 9060394	2.3
Kidney Ca, Nuclear grade 3 (OD04348)	77.4	Gastric Cancer 064005	0.0

Table AGO. Panel 2D

Tissue Name	Rel. Exp.(%) Ag047, Run 144771648	Rel. Exp.(%) Ag047, Run 152940364	Rel. Exp.(%) Ag2679, Run 158633803	Rel. Exp.(%) Ag2728, Run 158561830
Normal Colon	5.5	8.9	7.4	10.4
CC Well to Mod Diff (ODO3866)	1.4	0.2	0.2	1.3

CC Margin (ODO3866)	0.4	0.5	0.4	0.1
CC Gr.2 rectosigmoid (ODO3868)	0.4	0.1	1.7	0.2
CC Margin (ODO3868)	2.5	0.9	0.7	0.7
CC Mod Diff (ODO3920)	0.0	0.0	0.0	0.2
CC Margin (ODO3920)	0.7	0.9	1.1	0.3
CC Gr.2 ascend colon (ODO3921)	0.0	0.7	0.6	0.2
CC Margin (ODO3921)	0.2	0.2	0.6	0.8
CC from Partial Hepatectomy (ODO4309) Mets	0.6	0.4	1.1	1.0
Liver Margin (ODO4309)	11.1	13.6	30.4	21.2
Colon mets to lung (OD04451-01)	0.0	0.4	0.1	0.2
Lung Margin (OD04451-02)	1.3	1.0	0.5	0.7
Normal Prostate 6546-1	19.1	0.7	1.0	2.3
Prostate Cancer (OD04410)	5.8	2.9	4.7	3.4
Prostate Margin (OD04410)	4.7	5.7	7.2	4.9
Prostate Cancer (OD04720-01)	2.9	3.3	3.4	3.3
Prostate Margin (OD04720-02)	12.5	10.7	8.3	15.2
Normal Lung 061010	1.5	2.7	6.8	5.4
Lung Met to Muscle (ODO4286)	0.3	0.1	0.0	0.5
Muscle Margin (ODO4286)	0.3	0.2	0.3	0.6
Lung Malignant Cancer (OD03126)	0.4	0.2	0.8	1.5
Lung Margin (OD03126)	0.4	1.0	1.2	0.5
Lung Cancer (OD04404)	86.5	100.0	100.0	100.0

Lung Margin (OD04404)	18.3	3.3	2.2	3.0
Lung Cancer (OD04565)	100.0	52.1	62.0	77.9
Lung Margin (OD04565)	0.2	0.1	0.0	0.6
Lung Cancer (OD04237-01)	6.3	1.5	3.2	3.3
Lung Margin (OD04237-02)	1.4	0.5	0.5	0.6
Ocular Mel Met to Liver (ODO4310)	0.4	0.3	0.5	0.7
Liver Margin (ODO4310)	2.3	1.8	3.5	2.8
Melanoma Mets to Lung (OD04321)	0.0	0.3	0.4	1.6
Lung Margin (OD04321)	2.1	2.6	2.2	3.0
Normal Kidney	6.7	4.9	8.4	7.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	0.0	0.3	1.4
Kidney Margin (OD04338)	3.5	1.5	4.1	3.5
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	0.1	0.3	0.7
Kidney Margin (OD04339)	18.4	10.3	8.4	15.5
Kidney Ca, Clear cell type (OD04340)	0.0	0.8	0.7	1.1
Kidney Margin (OD04340)	6.5	4.4	4.5	6.5
Kidney Ca, Nuclear grade 3 (OD04348)	90.8	36.1	54.3	50.3
Kidney Margin (OD04348)	4.6	3.2	3.8	4.0
Kidney Cancer (OD04622-01)	2.7	1.8	4.1	3.5
Kidney Margin (OD04622-03)	2.0	0.2	0.3	1.1
Kidney Cancer (OD04450-01)	0.0	0.0	0.0	0.0
Kidney Margin (OD04450-03)	3.1	1.4	6.9	5.4
Kidney Cancer 8120607	1.5	0.3	0.5	1.8

Kidney Margin 8120608	3.4	0.8	1.3	2.2
Kidney Cancer 8120613	0.4	0.8	4.3	2.0
Kidney Margin 8120614	2.8	1.2	4.9	4.0
Kidney Cancer 9010320	76.8	39.0	36.3	52.9
Kidney Margin 9010321	10.4	5.1	4.9	4.3
Normal Uterus	0.0	0.2	0.3	0.0
Uterus Cancer 064011	0.9	0.0	0.3	1.9
Normal Thyroid	0.3	0.0	0.9	2.2
Thyroid Cancer 064010	0.0	0.0	0.0	0.0
Thyroid Cancer A302152	1.4	0.1	0.4	0.0
Thyroid Margin A302153	1.7	0.9	6.4	4.1
Normal Breast	20.3	5.7	13.0	20.9
Breast Cancer (OD04566)	0.3	0.1	0.5	0.3
Breast Cancer (OD04590-01)	2.4	2.2	1.1	1.9
Breast Cancer Mets (OD04590-03)	0.7	0.1	0.3	0.0
Breast Cancer Metastasis (OD04655-05)	6.9	3.3	7.7	9.3
Breast Cancer 064006	5.6	8.3	4.2	4.9
Breast Cancer 1024	47.0	19.3	30.1	23.5
Breast Cancer 9100266	14.9	9.3	15.7	21.8
Breast Margin 9100265	4.6	1.5	4.4	8.4
Breast Cancer A209073	48.3	12.9	28.3	40.9
Breast Margin A2090734	38.7	16.6	27.5	29.7
Normal Liver	0.2	0.1	0.0	0.2
Liver Cancer 064003	11.8	5.6	4.6	4.5
Liver Cancer 1025	4.5	1.6	4.2	2.8
Liver Cancer 1026	6.2	6.7	9.0	6.4

Liver Cancer 6004-T	15.6	3.3	3.5	2.8
Liver Tissue 6004-N	0.1	0.2	0.2	0.4
Liver Cancer 6005-T	14.6	8.0	8.7	7.9
Liver Tissue 6005-N	6.4	7.0	5.9	3.2
Normal Bladder	1.3	0.9	1.6	0.9
Bladder Cancer 1023	0.4	0.2	0.3	0.3
Bladder Cancer A302173	7.5	5.3	12.1	19.8
Bladder Cancer (OD04718-01)	27.7	23.0	35.6	41.5
Bladder Normal Adjacent (OD04718-03)	0.4	0.3	0.6	0.7
Normal Ovary	1.4	0.4	1.1	0.3
Ovarian Cancer 064008	1.3	0.5	1.3	1.3
Ovarian Cancer (OD04768-07)	0.4	0.2	0.0	0.0
Ovary Margin (OD04768-08)	1.0	0.8	0.6	1.6
Normal Stomach	4.3	2.0	0.3	3.1
Gastric Cancer 9060358	2.4	0.7	0.7	1.2
Stomach Margin 9060359	0.0	0.0	0.7	4.3
Gastric Cancer 9060395	0.9	2.0	1.9	2.4
Stomach Margin 9060394	0.2	0.3	1.1	0.2
Gastric Cancer 9060397	0.4	0.5	1.6	0.7
Stomach Margin 9060396	0.0	0.1	0.2	0.2
Gastric Cancer 064005	0.7	0.7	0.9	1.5

Table AGP. Panel 3D

Tissue Name	Rel. Exp.(%) Ag047, Run 158634002	Tissue Name	Rel. Exp.(%) Ag047, Run 158634002
Daoy- Medulloblastoma	0.9	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.3
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	6.7

D283 Med-Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megakaryoblast)	0.0
SNB-78- Glioma	100.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	79.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	1.4	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	2.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.1	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	4.5	JM1- pre-B-cell lymphoma	0.0
Cerebellum	2.9	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	3.8	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.1	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	1.0	769-P- Clear cell renal carcinoma	0.1
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.1
NCI-H82- Small cell lung cancer	7.6	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.8	G401- Wilms' tumor	6.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.2	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.1
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.6
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.0

Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.1
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	3.8
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinma (transitional cell)	10.3
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	59.9
NCI-SNU-5- Gastric carcinoma	0.7	MG-63- Osteosarcoma	0.5
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	4.2
NCI-SNU-16- Gastric carcinoma	0.2	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.3
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.6
RF-1- Gastric adenocarcinoma	0.3	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.1	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.2
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	1.1
RL95-2- Uterine carcinoma	4.0	SCC-15- Squamous cell carcinoma of tongue	0.3
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	7.0

Table AGQ. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag2975, Run 171818669	Tissue Name	Rel. Exp.(%) Ag2975, Run 171818669
-------------	--	-------------	--

Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.1
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	37.4
Primary Th2 rest	0.0	Small airway epithelium none	47.6
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	0.8	Coronary artery SMC rest	0.8
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.6
CD8 lymphocyte act	0.0	Astrocytes rest	51.8
Secondary CD8 lymphocyte rest	0.1	Astrocytes TNFalpha + IL-1beta	78.5
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	45.1
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	33.7
LAK cells IL-2	0.0	Liver cirrhosis	0.2
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.5
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	12.6
LAK cells IL-2+ IL-18	0.1	NCI-H292 IL-9	0.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	10.9
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	0.9
Two Way MLR 3 day	0.0	HPAEC none	1.2
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1	0.3

		beta	
Two Way MLR 7 day	0.1	Lung fibroblast none	1.5
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	0.0	Lung fibroblast IL-4	0.1
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.2
Ramos (B cell) none	0.0	Lung fibroblast IL-13	1.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.8
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	4.1
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	2.2
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.8
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	2.4
Dendritic cells none	0.0	Dermal fibroblast IL-4	5.4
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	9.3
Dendritic cells anti- CD40	0.1	Neutrophils TNFa+LPS	0.6
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	0.5
Macrophages rest	0.0	Lung	1.6
Macrophages LPS	0.0	Thymus	9.3
HUVEC none	0.0	Kidney	11.3
HUVEC starved	0.4		

Table AGR. Panel 4D

Tissue Name	Rel. Exp.(%) Ag047, Run 146087309	Rel. Exp.(%) Ag047, Run 151918119	Rel. Exp.(%) Ag047, Run 152893522	Rel. Exp.(%) Ag2679, Run 158535667	Rel. Exp.(%) Ag2728, Run 158562582	Rel. Exp.(%) Ag2975, Run 164314612	Rel. Exp.(%) Ag47b, Run 158664022
Secondary Th1 act	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Secondary Th2 act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Secondary Tr1 act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Secondary Th1 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Secondary Th2 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Secondary Tr1 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Th1 act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Th2 act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Tr1 act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Th1 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Th2 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Primary Tr1 rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CD45RA CD4 lymphocyt e act	0.1	0.2	0.1	0.3	0.3	0.0	0.3
CD45RO CD4 lymphocyt e act	0.0	0.0	0.0	0.0	0.0	0.0	0.1
CD8 lymphocyt e act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Secondary CD8 lymphocyt e rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Secondary CD8 lymphocyt e act	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CD4 lymphocyt e none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2ry Th1/Th2/T r1_anti- CD95 CH11	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LAK cells rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LAK cells IL-2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LAK cells IL-2+IL-	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1076

cells LPS							
Dendritic cells anti-CD40	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocytes rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocytes LPS	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Macrophages rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Macrophages LPS	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC starved	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IL-1beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IFN gamma	0.0	0.0	0.0	0.1	0.0	0.0	0.0
HUVEC TNF alpha + IFN gamma	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC TNF alpha + IL4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HUVEC IL-11	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung Microvascular EC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung Microvascular EC TNFalpha + IL-1beta	0.0	7.3	0.0	0.0	0.0	0.0	0.0
Microvascular Dermal EC none	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Microvascular Dermal EC TNFalpha	0.0	0.0	0.0	0.0	0.0	0.0	0.0

+ IL-1beta							
Bronchial epithelium TNFalpha + IL1beta	8.6	2.0	1.4	10.0	4.6	11.9	13.3
Small airway epithelium none	21.8	9.5	17.8	17.7	17.8	12.2	17.2
Small airway epithelium TNFalpha + IL-1beta	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Coronary artery SMC rest	0.2	0.2	0.3	0.4	0.4	0.2	0.2
Coronary artery SMC TNFalpha + IL-1beta	0.2	0.3	0.2	0.2	0.2	0.1	0.2
Astrocytes rest	23.3	19.3	20.4	30.8	20.7	16.6	21.8
Astrocytes TNFalpha + IL-1beta	27.0	25.0	25.3	27.9	22.7	19.3	28.5
KU-812 (Basophil) rest	0.0	0.0	0.0	0.0	0.0	0.0	0.0
KU-812 (Basophil) PMA/iono mycin	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CCD1106 (Keratinocytes) none	18.4	12.7	14.3	16.7	12.5	10.2	13.5
CCD1106 (Keratinocytes) TNFalpha + IL-1beta	4.2	1.4	2.0	7.1	3.5	5.1	8.4
Liver cirrhosis	0.3	0.2	0.2	0.3	0.3	0.2	0.4
Lupus kidney	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NCI-H292 none	0.4	0.2	0.4	0.3	0.6	0.1	0.2

NCI-H292 IL-4	9.3	4.8	7.5	8.9	8.2	5.3	8.7
NCI-H292 IL-9	0.6	0.3	0.6	0.7	0.7	0.5	0.4
NCI-H292 IL-13	4.0	5.1	3.9	4.5	4.2	2.9	3.3
NCI-H292 IFN gamma	0.2	0.2	0.1	0.3	0.2	0.2	0.2
HPAEC none	0.0	0.0	0.0	0.0	0.0	0.0	0.1
HPAEC TNF alpha + IL-1 beta	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lung fibroblast none	0.1	0.3	0.3	0.3	0.3	0.1	0.2
Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0	0.1	0.0	0.0	0.1	0.1
Lung fibroblast IL-4	0.3	0.1	0.2	0.3	0.1	0.2	0.3
Lung fibroblast IL-9	0.0	0.2	0.1	0.1	0.1	0.1	0.1
Lung fibroblast IL-13	0.1	0.1	0.1	0.2	0.2	0.0	0.3
Lung fibroblast IFN gamma	0.1	0.0	0.0	0.1	0.1	0.2	0.1
Dermal fibroblast CCD1070 rest	2.2	1.3	2.4	3.1	2.5	1.5	2.4
Dermal fibroblast CCD1070 TNF alpha	1.0	0.8	1.1	2.1	1.8	0.9	1.2
Dermal fibroblast CCD1070 IL-1 beta	0.3	0.3	0.6	0.7	0.6	0.4	0.6
Dermal fibroblast	0.6	0.6	1.1	1.5	1.1	0.6	0.7

IFN gamma							
Dermal fibroblast IL-4	1.8	1.2	2.0	2.5	1.5	1.2	1.4
IBD Colitis 2	0.1	0.1	0.0	0.0	0.0	0.0	0.0
IBD Crohn's	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Colon	0.0	0.0	0.1	0.2	0.2	0.1	0.2
Lung	0.3	0.3	0.2	0.6	0.3	0.2	0.2
Thymus	0.8	1.5	0.9	1.4	1.2	1.0	1.4
Kidney	3.0	2.8	3.6	4.5	3.3	2.0	2.6

CNS_neurodegeneration_v1.0 Summary: Ag047/Ag2679/Ag2728 This panel does not show differential expression of the NOV37 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1 for discussion of utility of this gene in the central nervous system.

Panel 1 Summary: Ag047/Ag332/Ag47b Multiple experiments with three different probe and primer sets produce results that are in excellent agreement, with highest expression of the NOV37 gene in a brain cancer cell line (CT=23-25). There is also significant expression in clusters of samples from melanoma, ovarian cancer, breast, lung, renal colon and brain cancer lines. Thus, expression of this gene may be associated with these forms of cancer and could potentially be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of these genes may be useful in the treatment of ovarian, breast, lung, renal, and brain cancer and melanoma.

In addition to significant expression in brain cancer cell lines, this gene is preferentially expressed in the brain. This expression profile suggests that this gene product may play a role in CNS processes. This gene encodes a homolog of a member of the neurestin family, Ten M2, and may play a role in neuronal regeneration. Thus, agents that induce the expression or activity of NOV37 may have utility as neuronal regeneration drugs. Such agents would have utility in neurodegenerative diseases, stroke, and neuronal trauma.

Among tissues with metabolic function, this gene shows consistent expression in thyroid, adult and fetal heart, liver and skeletal muscle. Thus, this gene product may be an antibody target for the treatment of metabolic and endocrine disease, including obesity and Types 1 and 2 diabetes. In addition, this gene is expressed at higher levels in adult liver than in fetal liver and may be useful for differentiating between the two sources of liver tissue.

References:

Otaki JM, Firestein S. Neurestin: putative transmembrane molecule implicated in neuronal development. Dev Biol 1999 Aug 1;212(1):165-81

We have cloned a novel cDNA encoding a putative transmembrane protein, neurestin, from the rat olfactory bulb. Neurestin was identified based on a sequence similar to that of the second extracellular loops of odorant receptors in the cysteine-rich CC box located immediately after EGF-like motifs. Neurestin shows homology to a neuregulin gene product, human gamma-heregulin, a Drosophila receptor-type pair-rule gene product, Odd Oz (Odz) / Ten(m), and Ten(a), suggesting a possible function in synapse formation and morphogenesis. Recently, a mouse neurestin homolog has independently been cloned as DOC4 from the NIH-3T3 cell line. Northern blot analysis showed that neurestin is highly expressed in the brain and also in other tissues at much lower levels. In situ hybridization studies showed that neurestin is expressed in many types of neurons, including pyramidal cells in the cerebral cortex and tufted cells in the olfactory bulb during development. In adults, neurestin is mainly expressed in olfactory and hippocampal granule cells, which are known to be generated throughout adulthood. Nonetheless, in adults the expression of neurestin was experimentally induced in external tufted cells during regeneration of olfactory sensory neurons. These results suggest a role for neurestin in neuronal development and regeneration in the central nervous system.

Panels 1.1/1.2/1.3D Summary: Ag047 The NOV37 gene is expressed in melanoma, ovarian cancer, breast, lung, renal and brain cancer lines with good concordance for 3 independent runs. Expression of this gene might be associated with these forms of cancer and therapeutic modulation of this gene might be of use in the treatment of these cancers. Please note that results from one experiment on Panel 1.3D with the probe/primer set Ag2975 are not included. The amp plot suggests that there were experimental difficulties with this run.

Panel 2.2 Summary: Ag2975 This gene appears to be expressed at a very low level in the samples used in this panel. Significant expression is only seen in lung, kidney and breast cancer samples. Expression of this gene might be associated with these forms of cancer and therapeutic modulation of this gene might be of use in the treatment of these cancers.

Panel 2D Summary: Ag047 The expression of the NOV37 gene was assessed in multiple runs on this panel, with excellent concordance between the runs. This gene appears to be expressed at a higher level in gastric, bladder, and 2 samples each of lung and kidney cancer when compared to the normal adjacent tissue. Thus, expression of this gene might be associated with these forms of cancer and therapeutic modulation of this gene might be of use in the treatment of these cancers.

Panel 3D Summary: Ag047 The NOV37 gene is expressed in squamous cell carcinoma, glioma, small cell lung cancer cell lines. Thus, expression of this gene might be associated with these cancers and therapeutic modulation of this gene might be of use in the treatment of these cancers.

5 **Panels 4D/4.1D Summary:** Ag047/Ag2679/Ag2728/Ag2975/Ag047b Multiple runs with different set of primers give very consistent expression data. Highest expression of the NOV37 transcript is found in small airway epithelium upon treatment with TNF- α and IL-1 (CT=25). This expression is significantly up-regulated when compared to untreated tissue (CT=28). Moderate expression of this transcript is also found in keratinocytes and astrocytes.

10 NOV37 encodes for a neurestin like molecule whose role in neuronal regeneration has been demonstrated. Therefore, the putative protein encoded by NOV37 may play an important role in the regeneration or repair mechanism of these tissues in inflammation. Thus, therapeutic modulation of the expression of this gene product may be beneficial for the treatment of inflammatory lung diseases such as bronchitis, chronic obstructive pulmonary disease,

15 emphysema. Therapeutics designed against this putative protein may also be useful for in the CNS for reducing inflammation, including inflammation that results from multiple sclerosis or stroke.

NOV38, NOV39a, and NOV39b

20 Expression of gene NOV38 and variants NOV39a and NOV39b was assessed using the primer-probe set Ag3753, described in Table AHA. Results of the RTQ-PCR runs are shown in Tables AHB and AHC.

Table AHA. Probe Name Ag3753

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cactgcagtaattcagctggta-3'	22	835	1135
Probe	TET-5'-agtatccagtcgcccatcccagtt-3'-TAMRA	25	797	1136
Reverse	5'-aggcgagaccattacgtagact-3'	22	769	1137

Table AHB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3753, Run 216707728	Tissue Name	Rel. Exp.(%) Ag3753, Run 216707728
Adipose	0.0	Renal ca. TK-10	1.0
Melanoma* Hs688(A).T	0.3	Bladder	0.0

Melanoma* Hs688(B).T	0.3	Gastric ca. (liver met.) NCI-N87	0.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.2	Colon ca.* (SW480 met) SW620	2.6
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.1	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.1	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV- 3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.1	Colon Pool	0.0
Ovarian ca. OVCAR-5	100.0	Small Intestine Pool	0.0
Ovarian ca. IGROV- 1	0.1	Stomach Pool	0.0
Ovarian ca. OVCAR-8	2.1	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.1	Heart Pool	0.9
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.5	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.2
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.2
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.2	CNS cancer (astro) SF- 539	1.1
Lung ca. LX-1	1.9	CNS cancer (astro) SNB-75	0.4

Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	0.2
Lung ca. SHP-77	0.9	CNS cancer (glio) SF- 295	3.1
Lung ca. A549	0.1	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.1
Lung ca. NCI-H23	0.2	Brain (fetal)	0.0
Lung ca. NCI-H460	0.1	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.2	Brain (Substantia nigra) Pool	0.0
Liver	1.2	Brain (Thalamus) Pool	0.0
Fetal Liver	0.9	Brain (whole)	0.1
Liver ca. HepG2	2.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.1	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table AHC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3753, Run 172209242	Tissue Name	Rel. Exp.(%) Ag3753, Run 172209242
Secondary Th1 act	0.5	HUVEC IL-1beta	0.5
Secondary Th2 act	2.1	HUVEC IFN gamma	1.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	1.2	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.2	HUVEC IL-11	0.0
Secondary Tr1 rest	1.0	Lung Microvascular EC none	0.9
Primary Th1 act	0.5	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.2	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.5	Bronchial epithelium TNFalpha + IL1beta	1.2
Primary Th2 rest	0.0	Small airway epithelium	0.0

		none	
Primary Tr1 rest	0.5	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	1.7	Coronary artery SMC rest	2.6
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	1.3
CD8 lymphocyte act	0.0	Astrocytes rest	1.2
Secondary CD8 lymphocyte rest	0.2	Astrocytes TNFalpha + IL-1beta	0.4
Secondary CD8 lymphocyte act	3.6	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.5	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	2.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	39.5
LAK cells IL-2+IL-12	1.3	NCI-H292 none	0.3
LAK cells IL-2+IFN gamma	1.0	NCI-H292 IL-4	1.1
LAK cells IL-2+ IL-18	0.5	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	3.8	NCI-H292 IL-13	0.8
NK Cells IL-2 rest	0.3	NCI-H292 IFN gamma	0.3
Two Way MLR 3 day	0.5	HPAEC none	0.0
Two Way MLR 5 day	2.6	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	1.0	Lung fibroblast none	15.1
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	1.2
PBMC PWM	0.0	Lung fibroblast IL-4	0.8
PBMC PHA-L	0.5	Lung fibroblast IL-9	0.0
Ramos (B cell) none	12.8	Lung fibroblast IL-13	0.4
Ramos (B cell) ionomycin	100.0	Lung fibroblast IFN gamma	0.8
B lymphocytes PWM	0.5	Dermal fibroblast CCD1070 rest	3.7
B lymphocytes CD40L and IL-4	0.9	Dermal fibroblast CCD1070 TNF alpha	2.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	2.2
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0

Dendritic cells none	0.9	Dermal fibroblast IL-4	4.3
Dendritic cells LPS	6.2	Dermal Fibroblasts rest	0.9
Dendritic cells anti-CD40	3.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.5	Colon	0.8
Macrophages rest	3.7	Lung	0.0
Macrophages LPS	1.6	Thymus	1.1
HUVEC none	0.0	Kidney	15.3
HUVEC starved	0.0		

General_screening_panel_v1.4 Summary: Ag3753 Results from one experiment with the NOV38 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

- 5 **Panel 4.1D Summary:** Ag3753 Highest expression of the NOV38 transcript is found in Ramos B cell line activated with PMA and ionomycin (CT=29). However, expression is not seen in primary activated B cells. Therefore, expression of this gene could potentially be used as a marker for activated B lymphoma. This gene is also expressed at lower levels in liver cirrhosis, lung fibroblasts and kidney. This transcript encodes for a molecule that belongs to
- 10 the activin family, a member of the TGF beta superfamily. These factors influence growth and differentiation and are present in many cells and tissues. Therefore, therapeutics using the protein encoded by NOV38 could be important for the normal homeostasis of these tissues.

NOV40

- 15 Expression of gene NOV40 was assessed using the primer-probe set Ag2907, described in Table AIA. Results of the RTQ-PCR runs are shown in Tables AIB, AIC, AID and AIE.

Table AIA. Probe Name Ag2907

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggctcattcgaaactactggta-3'	22	773	1138
Probe	TET-5'-tggaatttcctcgcccactcttacct-3'-TAMRA	26	797	1139
Reverse	5'-ggttgacaggtttgcagtagag-3'	22	844	1140

Table AIB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2907, Run	Rel. Exp.(%) Ag2907, Run	Tissue Name	Rel. Exp.(%) Ag2907, Run	Rel. Exp.(%) Ag2907, Run
-------------	-----------------------------	-----------------------------	-------------	-----------------------------	-----------------------------

	157283423	165701505		157283423	165701505
Liver adenocarcinoma	0.0	0.0	Kidney (fetal)	0.0	0.0
Pancreas	0.0	0.0	Renal ca. 786-0	0.0	0.0
Pancreatic ca. CAPAN 2	0.0	0.0	Renal ca. A498	0.0	0.0
Adrenal gland	0.0	0.0	Renal ca. RXF 393	0.0	0.0
Thyroid	0.0	0.0	Renal ca. ACHN	0.0	0.0
Salivary gland	0.0	0.0	Renal ca. UO-31	0.0	0.0
Pituitary gland	100.0	100.0	Renal ca. TK-10	2.0	28.3
Brain (fetal)	0.0	0.0	Liver	0.0	0.0
Brain (whole)	0.0	14.3	Liver (fetal)	0.0	0.0
Brain (amygdala)	0.0	0.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	0.0	11.9	Lung	0.0	0.0
Brain (hippocampus)	1.9	0.0	Lung (fetal)	0.0	0.0
Brain (substantia nigra)	0.0	0.0	Lung ca. (small cell) LX-1	0.0	0.0
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	1.1	0.0
Cerebral Cortex	2.0	0.0	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	1.3	0.0	Lung ca. (large cell) NCI-H460	0.0	0.0
glio/astro U87-MG	0.0	12.2	Lung ca. (non-sm. cell) A549	0.0	0.0
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.0	0.0
neuro*; met SK-N-AS	1.0	0.0	Lung ca. (non-s.cl) NCI-H522	0.0	0.0
astrocytoma SF-539	1.4	0.0	Lung ca. (squam.) SW	0.9	0.0

			900		
astrocytoma SNB-75	1.0	0.0	Lung ca. (squam.) NCI-H596	0.0	0.0
glioma SNB-19	1.8	0.0	Mammary gland	0.0	0.0
glioma U251	0.8	0.0	Breast ca.* (pl.ef) MCF-7	1.4	2.5
glioma SF-295	1.0	0.0	Breast ca.* (pl.ef) MDA-MB-231	2.6	0.0
Heart (fetal)	0.0	0.0	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	0.0	0.0	Breast ca. BT-549	3.3	0.0
Skeletal muscle (fetal)	1.0	0.0	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.0	0.0	Ovary	0.0	0.0
Bone marrow	0.0	0.0	Ovarian ca. OVCAR-3	2.9	0.0
Thymus	0.0	0.0	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.9	0.0	Ovarian ca. OVCAR-5	3.0	10.9
Lymph node	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.0	0.0	Ovarian ca. IGROV-1	0.0	0.0
Stomach	0.0	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0	0.0
Small intestine	0.0	0.0	Uterus	0.0	0.0
Colon ca. SW480	0.0	0.0	Placenta	0.9	0.0
Colon ca.* SW620(SW480 met)	0.0	0.0	Prostate	0.0	0.0
Colon ca. HT29	1.0	0.0	Prostate ca.* (bone met)PC-3	1.3	0.0
Colon ca. HCT-116	0.0	0.0	Testis	16.6	9.5
Colon ca. CaCo-2	1.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	0.0	11.0	Melanoma* (met) Hs688(B).T	0.0	0.0

Colon ca. HCC-2998	2.6	0.0	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	1.2	13.3	Melanoma M14	0.0	0.0
Bladder	0.0	6.7	Melanoma LOX IMVI	0.0	0.0
Trachea	0.0	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	0.0	0.0	Adipose	1.5	0.0

Table AIC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2907, Run 157284121	Tissue Name	Rel. Exp.(%) Ag2907, Run 157284121
Normal Colon	4.5	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	2.1	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	2.3	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	1.7	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	4.8
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	1.5
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.0
CC Margin (ODO3921)	6.8	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	15.2
Colon mets to lung (OD04451-01)	0.0	Normal Breast	1.7
Lung Margin (OD04451-02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	2.2
Prostate Margin	0.0	Breast Cancer	2.8

(OD04410)		Metastasis (OD04655-05)	
Prostate Cancer (OD04720-01)	100.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.0
Normal Lung 061010	7.2	Breast Cancer 9100266	2.4
Lung Met to Muscle (ODO4286)	0.6	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	3.4	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	3.5
Lung Margin (OD04404)	4.1	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237- 01)	0.0	Liver Tissue 6004-N	0.5
Lung Margin (OD04237- 02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	22.1
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	2.4
Kidney Ca, Nuclear grade 2 (OD04338)	2.0	Bladder Normal Adjacent (OD04718- 03)	0.0
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	2.0	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	14.4
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	2.4	Normal Stomach	0.0

Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	0.0	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	2.1	Gastric Cancer 064005	2.9

Table AID. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2907, Run 164633936	Tissue Name	Rel. Exp.(%) Ag2907, Run 164633936
Daoy- Medulloblastoma	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	49.7
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	15.7	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	49.7
Cerebellum	0.0	JM1- pre-B-cell lymphoma	0.0
Cerebellum	0.0	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	94.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0

DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	27.2
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	16.4	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.0
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinoma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	17.6
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinoma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0

NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	37.6
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	100.0

Table AIE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2907, Run 157284733	Tissue Name	Rel. Exp.(%) Ag2907, Run 157284733
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	12.3	Lung Microvascular EC TNFalpha + IL-1beta	10.4
Primary Th2 act	0.0	Microvascular Dermal EC none	11.2
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	11.7	Bronchial epithelium TNFalpha + IL1beta	12.9
Primary Th2 rest	0.0	Small airway epithelium none	24.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0

CD45RO CD4 lymphocyte act	6.1	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	12.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	4.5	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	13.4	Liver cirrhosis	35.1
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	33.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	36.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	23.3
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	21.2
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	11.6
Two Way MLR 5 day	0.0	HPAEC none	13.3
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	11.7	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	13.1

Monocytes rest	0.0	IBD Crohn's	6.9
Monocytes LPS	26.1	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	16.0
HUVEC none	0.0	Kidney	13.1
HUVEC starved	0.0		

Panel 1.3D Summary: Ag2907 Results from two experiments with the same probe/primer set are in good agreement. Expression of the NOV40 gene is highest in a sample derived from pituitary tissue with little to no expression detected in any other tissue. Thus, expression of this gene could be used to distinguish pituitary gland from the other samples on this panel.

The protein encoded for by this gene is most homologous to a glucuronosyltransferase, normally found in liver. UDP glycosyltransferases (UGT) are a superfamily of enzymes that catalyze the addition of the glycosyl group from a UTP-sugar to a small hydrophobic molecule. Glucuronosyltransferases are membrane-bound microsomal enzymes that catalyze the transfer of glucuronic acid to a wide variety of exogenous and endogenous lipophilic substrates. These enzymes are of major importance in the detoxification and subsequent elimination of xenobiotics such as drugs and carcinogens. The pituitary plays a major role in the physiology of many different systems in the body. Therefore, this gene may play an essential role in maintaining proper function of the pituitary gland and many of its secreted peptides. Furthermore, therapeutic modulation of the activity of this gene or its protein product using small molecule drugs may be useful for the treatment of diabetes and diabetes as well as growth, reproductive, and endocrine disorders.

Panel 2D Summary: Ag2907 Expression of the NOV40 gene is highest and almost exclusive to a sample derived from a prostate cancer (CT = 31.7). Thus, the expression of this gene could be used to distinguish prostate cancer from the other samples in the panel. Moreover, therapeutic modulation of the activity of this gene or its protein product, through the use of small molecule drugs, protein therapeutics or antibodies, might be of benefit in the treatment of prostate cancer.

Panel 3D Summary: Ag2907 Expression of the NOV40 gene is highest in a sample derived from a squamous cell carcinoma cell line (CT = 33.8). Thus, the expression of this gene could be used to distinguish this sample from the other samples in the panel.

Panel 4D Summary: Ag2907 Expression of the NOV40 gene is detected at a very low level in small airway epithelium treated with the inflammatory cytokines TNF- α and IL-1 β (CT = 34.2). Thus, expression of this gene may be a marker of inflammation in the lung.

NOV41a and NOV41b

5 Expression of gene NOV41a and variant NOV41b was assessed using the primer-probe sets Ag1361 and Ag2953, described in Tables AJA and AJB. Results of the RTQ-PCR runs are shown in Tables AJC, AJD and AJE.

Table AJA. Probe Name Ag1361

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctggtcaggtacctggatgtta-3'	22	1438	1141
Probe	TET-5'-tccatcaatgaagagcttcattcg-3'-TAMRA	26	1480	1142
Reverse	5'-cagcctttaagtgatccatcag-3'	22	1507	1143

10

Table AJB. Probe Name Ag2953

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctggtcaggtacctggatgtta-3'	22	1438	1144
Probe	TET-5'-tccatcaatgaagagcttcattcg-3'-TAMRA	26	1480	1145
Reverse	5'-cagcctttaagtgatccatcag-3'	22	1507	1146

Table AJC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1361, Run 152953143	Tissue Name	Rel. Exp.(%) Ag1361, Run 152953143
Liver adenocarcinoma	0.0	Kidney (fetal)	2.1
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	1.6	Renal ca. UO-31	0.0
Pituitary gland	0.1	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.2	Liver (fetal)	0.0
Brain (amygdala)	0.7	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	1.0	Lung (fetal)	0.0

Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.2	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.1
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.2
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.5
Thymus	0.4	Ovarian ca. OVCAR- 4	0.0
Spleen	0.0	Ovarian ca. OVCAR- 5	0.7
Lymph node	0.0	Ovarian ca. OVCAR- 8	0.0
Colorectal	0.0	Ovarian ca. IGROV- 1	0.0
Stomach	100.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.*	0.0	Prostate	0.0

SW620(SW480 met)			
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	2.1
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.4	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.4	Melanoma* (met) SK-MEL-5	0.0
Kidney	12.8	Adipose	0.0

Table AJD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag1361, Run 152953321	Tissue Name	Rel. Exp.(%) Ag1361, Run 152953321
Normal Colon	0.2	Kidney Margin 8120608	2.1
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.5
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	2.6
CC Gr.2 rectosigmoid (ODO3868)	1.3	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	2.9
CC Mod Diff (ODO3920)	0.6	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.2
CC Gr.2 ascend colon (ODO3921)	0.2	Normal Thyroid	0.0
CC Margin (ODO3921)	0.0	Thyroid Cancer	0.0

		064010	
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.0
Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	0.0
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0
Prostate Margin (OD04720-02)	0.1	Breast Cancer 1024	0.3
Normal Lung 061010	0.3	Breast Cancer 9100266	0.0
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.1
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.1
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.1	Liver Cancer 064003	0.0

Lung Margin (OD04404)	0.1	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.2
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.4
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.1
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	0.1
Normal Kidney	23.3	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.7	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	9.2	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	0.2
Kidney Margin (OD04339)	28.3	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	45.4	Normal Stomach	100.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	5.3

Kidney Margin (OD04348)	18.6	Stomach Margin 9060359	78.5
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.3
Kidney Margin (OD04622-03)	3.4	Stomach Margin 9060394	31.6
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.2
Kidney Margin (OD04450-03)	25.7	Stomach Margin 9060396	29.5
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	4.5

Table AJE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1361, Run 152953376	Rel. Exp.(%) Ag2953, Run 164306345	Tissue Name	Rel. Exp.(%) Ag1361, Run 152953376	Rel. Exp.(%) Ag2953, Run 164306345
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microsvascular Dermal EC TNFalpha + IL-	0.0	0.0

			lbeta		
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	0.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	0.0
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	0.0
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL-1beta	0.1	0.2
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0	0.0
LAK cells IL-2	0.0	0.0	Liver cirrhosis	0.0	0.0
LAK cells IL-2+IL-12	0.0	0.0	Lupus kidney	1.2	0.8
LAK cells IL-2+IFN gamma	0.0	0.0	NCI-H292 none	0.2	0.0
LAK cells IL-2+IL-18	0.0	0.0	NCI-H292 IL-4	0.0	0.0
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-9	0.2	0.0
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	0.0	0.0
Two Way MLR 3 day	0.0	0.0	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 5 day	0.0	0.0	HPAEC none	0.0	0.0

Two Way MLR 7 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast none	0.0	0.0
PBMC PWM	0.0	0.2	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
B lymphocytes PWM	0.0	0.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes CD40L and IL-4	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	IBD Colitis 2	0.0	0.0
Monocytes rest	0.0	0.0	IBD Crohn's	0.0	0.0
Monocytes LPS	0.0	0.0	Colon	0.0	0.0
Macrophages rest	0.0	0.0	Lung	0.0	0.0
Macrophages LPS	0.0	0.0	Thymus	100.0	100.0
HUVEC none	0.0	0.0	Kidney	0.0	0.2
HUVEC starved	0.0	0.0			

Panel 1.3D Summary: Ag1361 Expression of the NOV41a gene is restricted to stomach (CT value = 29.9) and kidney (CT value = 32.9) tissue. This observation is consistent with the identification of this gene as a sodium/hydrogen ion exchanger because the function of both of these tissues requires sodium/hydrogen ion exchange activity. The inhibition of the NOV41A protein activity, through the use of antibodies or small molecule drugs, might be of use in the treatment of kidney or gastric diseases related to the function of a sodium/hydrogen ion exchanger. For example, the activity of this gene may be related to over-production of

stomach acid leading to acid reflux disease or peptic ulcer. Results from a second experiment with the probe and primer set Ag2953 are not included. The amp plot indicates that there is a potential problem in one of the sample wells.

Panel 2D Summary: Ag1361 Consistent with what was observed in Panel 1.3D, expression of the NOV41a gene in panel 2D is restricted to both normal kidney and stomach adjacent to tumor tissue. Interestingly, expression of the gene is absent in 4/4 gastric tumors and 10/10 kidney cancers when compared to the normal adjacent tissue controls. Thus, the expression of this gene appears to be a consistent trait of the non-neoplastic kidney and stomach. Therefore the absence of expression of this gene could be used as a diagnostic marker for kidney or gastric cancer. In addition, the replacement of this gene, potentially through the direct application of the protein or using gene replacement therapy, could be of use in the treatment of kidney or gastric cancer. Na⁺/H⁺ exchangers have previously been implicated in modulation of cellular adhesion and tumor invasion (Refs. 1 and 2).

References:

1. Denker S.P., Huang D.C., Orlowski J., Furthmayr H., Barber D.L. (2000) Direct binding of the Na⁺-H exchanger NHE1 to ERM proteins regulates the cortical cytoskeleton and cell shape independently of H⁺ translocation. *Mol. Cell* 6: 1425-1436.

The association of actin filaments with the plasma membrane maintains cell shape and adhesion. Here, we show that the plasma membrane ion exchanger NHE1 acts as an anchor for actin filaments to control the integrity of the cortical cytoskeleton. This occurs through a previously unrecognized structural link between NHE1 and the actin binding proteins ezrin, radixin, and moesin (ERM). NHE1 and ERM proteins associate directly and colocalize in lamellipodia. Fibroblasts expressing NHE1 with mutations that disrupt ERM binding, but not ion translocation, have impaired organization of focal adhesions and actin stress fibers, and an irregular cell shape. We propose a structural role for NHE1 in regulating the cortical cytoskeleton that is independent of its function as an ion exchanger.

PMID: 11163215

2. Reshkin S.J., Bellizzi A., Albarani V., Guerra L., Tommasino M., Paradiso A., Casavola V. (2000) Phosphoinositide 3-kinase is involved in the tumor-specific activation of human breast cancer cell Na⁽⁺⁾/H⁽⁺⁾ exchange, motility, and invasion induced by serum deprivation. *J. Biol. Chem.* 275: 5361-5369.

Whereas the tumor acidic extracellular pH plays a crucial role in the invasive process, the mechanism(s) behind this acidification, especially in low nutrient conditions, are unclear. The regulation of the Na⁽⁺⁾/H⁽⁺⁾ exchanger (NHE) and invasion by serum deprivation were

studied in a series of breast epithelial cell lines representing progression from non-tumor to highly metastatic cells. Whereas serum deprivation reduced lactate production in all three cells lines, it inhibited NHE activity in the non-tumor cells and stimulated it in the tumor cells with a larger stimulation in the metastatic cells. The stimulation of NHE in the tumor cell lines was the result of an increased affinity of the internal H(+) regulatory site of the NHE without changes in sodium kinetics or expression. Serum deprivation conferred increased cell motility and invasive ability that were abrogated by specific inhibition of the NHE. Inhibition of phosphoinositide 3-kinase by overexpression of a dominant-negative mutant or wortmannin incubation inhibited NHE activity and invasion in serum replete conditions while potentiating the serum deprivation-dependent activation of the NHE and invasion. These results indicate that the up-regulation of the NHE by a phosphoinositide 3-kinase-dependent mechanism plays an essential role in increased tumor cell invasion induced by serum deprivation.

PMID: 10681510

Panel 4D Summary: Ag1361/Ag2953 Two experiments with the same probe and primer sets produce results that are in excellent agreement. The NOV41a transcript is expressed in the thymus in Panel 4D (CT = 28.6), but not in Panel 1.3D (CT = 38). The NOV41A gene encodes a putative ion exchange molecule and may therefore be important in signal transduction in the thymus. Antibodies against the protein encoded for by the NOV41A gene may be used to identify thymic tissue. Additionally, small molecule or antibody therapeutics designed against this putative ion exchanger could disrupt T cell development in the thymus and serve an immunosuppressive function that could be important for tissue transplant.

NOV42a and NOV42b

Expression of gene NOV42a and variant NOV42b was assessed using the primer-probe set Ag3002, described in Table AKA. Results of the RTQ-PCR runs are shown in Tables AKB and AKC.

Table AKA. Probe Name Ag3002

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5' - agaccctccatgtggtcatt - 3'	20	1188	1147
Probe	TET-5' - tcacaggaacagctacaaagaaccca - 3' - TAMRA	26	1211	1148
Reverse	5' - caggaccatctggagaagct - 3'	20	1245	1149

[illegible]1106

Skeletal muscle (fetal)	6.5	Breast ca. MDA-N	0.0
Skeletal muscle	20.6	Ovary	12.5
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	13.7	Ovarian ca. OVCAR-4	0.0
Spleen	71.7	Ovarian ca. OVCAR-5	21.6
Lymph node	24.8	Ovarian ca. OVCAR-8	15.2
Colorectal	3.8	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	33.2	Uterus	27.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	4.2	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	2.4	Testis	7.7
Colon ca. CaCo-2	4.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	13.8

Table AKC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3002, Run 164043126	Tissue Name	Rel. Exp.(%) Ag3002, Run 164043126
Secondary Th1 act	10.2	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0

Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	14.2	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	6.2	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	10.8
LAK cells IL-2+IFN gamma	23.3	NCI-H292 none	10.8
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	30.6	Lung fibroblast none	0.0
PBMC PWM	24.8	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	13.5	Lung fibroblast IL-4	0.0

Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	100.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	99.3	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	22.4	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	36.3	IBD Colitis 2	0.0
Monocytes rest	76.8	IBD Crohn's	22.2
Monocytes LPS	0.0	Colon	67.8
Macrophages rest	0.0	Lung	22.7
Macrophages LPS	0.0	Thymus	12.2
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Panel 1.3D Summary: Ag3002 Significant expression of the NOV42a gene is limited to the mammary gland and the spleen (CTs=33-34). Thus, expression of this gene could be used to differentiate these samples from other samples on this panel and as a marker for these tissues.

Panel 4D Summary: Ag3002 Significant expression of the NOV42a gene is limited to activated B cells, an eosinophil cell line, and monocytes (CTs=33-35). Thus, this transcript could be used as a marker for phagocytic cell types.

NOV43

Expression of gene NOV43 was assessed using the primer-probe set Ag2987, described in Table ALA. Results of the RTQ-PCR runs are shown in Tables ALB, ALC, ALD and ALE.

Table ALA. Probe Name Ag2987

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctacctcaccatctgctttctg-3'	22	860	1150
Probe	TET-5'-tttctcaggactgccagctcttgatg-3'-	26	883	1151

	TAMRA			
Reverse	5'-tccatatcttgtaggccacact-3'	22	916	1152

Table ALB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2987, Run 211008970	Tissue Name	Rel. Exp.(%) Ag2987, Run 211008970
AD 1 Hippo	31.2	Control (Path) 3 Temporal Ctx	14.7
AD 2 Hippo	29.9	Control (Path) 4 Temporal Ctx	43.8
AD 3 Hippo	37.4	AD 1 Occipital Ctx	44.1
AD 4 Hippo	10.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	87.7	AD 3 Occipital Ctx	27.7
AD 6 Hippo	85.3	AD 4 Occipital Ctx	37.4
Control 2 Hippo	14.5	AD 5 Occipital Ctx	17.0
Control 4 Hippo	13.4	AD 6 Occipital Ctx	12.5
Control (Path) 3 Hippo	20.7	Control 1 Occipital Ctx	9.8
AD 1 Temporal Ctx	49.0	Control 2 Occipital Ctx	14.6
AD 2 Temporal Ctx	35.6	Control 3 Occipital Ctx	29.7
AD 3 Temporal Ctx	29.9	Control 4 Occipital Ctx	9.4
AD 4 Temporal Ctx	31.0	Control (Path) 1 Occipital Ctx	44.4
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	13.2
AD 5 SupTemporal Ctx	64.2	Control (Path) 3 Occipital Ctx	5.8
AD 6 Inf Temporal Ctx	69.3	Control (Path) 4 Occipital Ctx	29.9
AD 6 Sup Temporal Ctx	96.6	Control 1 Parietal Ctx	19.3
Control 1 Temporal Ctx	17.0	Control 2 Parietal Ctx	67.8
Control 2 Temporal Ctx	10.6	Control 3 Parietal Ctx	23.7
Control 3 Temporal Ctx	26.4	Control (Path) 1 Parietal Ctx	23.0
Control 4 Temporal Ctx	11.0	Control (Path) 2 Parietal Ctx	36.1
Control (Path) 1 Temporal Ctx	35.6	Control (Path) 3 Parietal Ctx	16.2

Control (Path) 2 Temporal Ctx	35.4	Control (Path) 4 Parietal Ctx	45.4
----------------------------------	------	----------------------------------	------

Table ALC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2987, Run 166232814	Tissue Name	Rel. Exp.(%) Ag2987, Run 166232814
Liver adenocarcinoma	5.9	Kidney (fetal)	0.0
Pancreas	6.4	Renal ca. 786-0	13.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	13.6
Adrenal gland	18.7	Renal ca. RXF 393	28.1
Thyroid	5.6	Renal ca. ACHN	0.0
Salivary gland	40.9	Renal ca. UO-31	0.0
Pituitary gland	23.5	Renal ca. TK-10	5.4
Brain (fetal)	67.4	Liver	0.0
Brain (whole)	84.7	Liver (fetal)	4.8
Brain (amygdala)	39.5	Liver ca. (hepatoblast) HepG2	7.0
Brain (cerebellum)	100.0	Lung	0.0
Brain (hippocampus)	21.8	Lung (fetal)	2.7
Brain (substantia nigra)	16.6	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	96.6	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	39.8	Lung ca. (s.cell var.) SHP-77	10.3
Spinal cord	41.2	Lung ca. (large cell)NCI-H460	13.0
glio/astro U87-MG	5.6	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	51.8	Lung ca. (non-s.cell) NCI-H23	6.4
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	10.4
neuro*; met SK-N-AS	8.4	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	4.3	Lung ca. (squam.) SW 900	34.2
astrocytoma SNB-75	10.4	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	28.5	Mammary gland	0.0
glioma U251	10.2	Breast ca.* (pl.ef) MCF-7	13.0
glioma SF-295	76.8	Breast ca.* (pl.ef)	0.0

		MDA-MB-231	
Heart (fetal)	5.1	Breast ca.* (pl.ef) T47D	0.0
Heart	10.5	Breast ca. BT-549	15.8
Skeletal muscle (fetal)	10.8	Breast ca. MDA-N	0.0
Skeletal muscle	15.7	Ovary	7.3
Bone marrow	5.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	5.8
Spleen	11.3	Ovarian ca. OVCAR-5	38.2
Lymph node	42.3	Ovarian ca. OVCAR-8	31.4
Colorectal	8.4	Ovarian ca. IGROV-1	10.2
Stomach	10.8	Ovarian ca.* (ascites) SK-OV-3	54.0
Small intestine	34.9	Uterus	25.7
Colon ca. SW480	0.0	Placenta	11.7
Colon ca.* SW620(SW480 met)	0.0	Prostate	18.9
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	9.1
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	4.8	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	11.9
Gastric ca.* (liver met) NCI-N87	31.2	Melanoma M14	0.0
Bladder	25.2	Melanoma LOX IMVI	5.2
Trachea	10.6	Melanoma* (met) SK-MEL-5	7.0
Kidney	0.0	Adipose	27.4

Table ALD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2987, Run 164314632	Tissue Name	Rel. Exp.(%) Ag2987, Run 164314632
Secondary Th1 act	8.1	HUVEC IL-1beta	1.9
Secondary Th2 act	10.3	HUVEC IFN gamma	17.4

Secondary Tr1 act	2.2	HUVEC TNF alpha + IFN gamma	9.6
Secondary Th1 rest	8.8	HUVEC TNF alpha + IL4	4.1
Secondary Th2 rest	10.0	HUVEC IL-11	9.3
Secondary Tr1 rest	25.0	Lung Microvascular EC none	31.9
Primary Th1 act	4.0	Lung Microvascular EC TNFalpha + IL-1beta	22.8
Primary Th2 act	1.9	Microvascular Dermal EC none	25.9
Primary Tr1 act	4.5	Microvascular Dermal EC TNFalpha + IL-1beta	7.3
Primary Th1 rest	40.1	Bronchial epithelium TNFalpha + IL1beta	33.4
Primary Th2 rest	39.5	Small airway epithelium none	10.6
Primary Tr1 rest	74.7	Small airway epithelium TNFalpha + IL-1beta	77.4
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	15.0
CD45RO CD4 lymphocyte act	15.4	Coronary artery SMC TNFalpha + IL-1beta	3.0
CD8 lymphocyte act	6.1	Astrocytes rest	4.6
Secondary CD8 lymphocyte rest	16.3	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	1.8	KU-812 (Basophil) rest	1.7
CD4 lymphocyte none	10.0	KU-812 (Basophil) PMA/ionomycin	20.9
2ry Th1/Th2/Tr1_anti-CD95 CH11	13.7	CCD1106 (Keratinocytes) none	8.4
LAK cells rest	33.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	7.9
LAK cells IL-2	14.8	Liver cirrhosis	18.6
LAK cells IL-2+IL-12	14.9	Lupus kidney	15.3
LAK cells IL-2+IFN gamma	28.9	NCI-H292 none	15.0
LAK cells IL-2+ IL-18	14.4	NCI-H292 IL-4	4.6
LAK cells PMA/ionomycin	4.5	NCI-H292 IL-9	13.2
NK Cells IL-2 rest	11.3	NCI-H292 IL-13	4.7
Two Way MLR 3 day	32.3	NCI-H292 IFN gamma	2.8
Two Way MLR 5 day	2.4	HPAEC none	8.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	10.2

PBMC rest	4.5	Lung fibroblast none	35.6
PBMC PWM	18.8	Lung fibroblast TNF alpha + IL-1 beta	16.3
PBMC PHA-L	20.2	Lung fibroblast IL-4	65.1
Ramos (B cell) none	1.4	Lung fibroblast IL-9	55.5
Ramos (B cell) ionomycin	5.6	Lung fibroblast IL-13	57.4
B lymphocytes PWM	19.2	Lung fibroblast IFN gamma	100.0
B lymphocytes CD40L and IL-4	11.7	Dermal fibroblast CCD1070 rest	9.7
EOL-1 dbcAMP	59.9	Dermal fibroblast CCD1070 TNF alpha	29.7
EOL-1 dbcAMP PMA/ionomycin	11.9	Dermal fibroblast CCD1070 IL-1 beta	3.2
Dendritic cells none	48.6	Dermal fibroblast IFN gamma	11.4
Dendritic cells LPS	17.7	Dermal fibroblast IL-4	24.3
Dendritic cells anti-CD40	29.5	IBD Colitis 2	6.7
Monocytes rest	26.4	IBD Crohn's	4.5
Monocytes LPS	19.5	Colon	20.4
Macrophages rest	28.1	Lung	14.8
Macrophages LPS	8.9	Thymus	52.5
HUVEC none	9.9	Kidney	74.7
HUVEC starved	21.2		

Table ALE. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag2987, Run 171670053	Tissue Name	Rel. Exp.(%) Ag2987, Run 171670053
BA4 Control	21.3	BA17 PSP	16.8
BA4 Control2	15.2	BA17 PSP2	4.4
BA4 Alzheimer's2	27.5	Sub Nigra Control	39.8
BA4 Parkinson's	100.0	Sub Nigra Control2	13.9
BA4 Parkinson's2	66.9	Sub Nigra Alzheimer's2	34.6
BA4 Huntington's	33.2	Sub Nigra Parkinson's2	51.4
BA4 Huntington's2	31.4	Sub Nigra Huntington's	69.3
BA4 PSP	3.9	Sub Nigra Huntington's2	22.7
BA4 PSP2	5.1	Sub Nigra PSP2	17.2

BA4 Depression	31.6	Sub Nigra Depression	23.0
BA4 Depression2	23.3	Sub Nigra Depression2	17.1
BA7 Control	39.5	Glob Palladus Control	23.7
BA7 Control2	6.0	Glob Palladus Control2	6.3
BA7 Alzheimer's2	43.2	Glob Palladus Alzheimer's	17.6
BA7 Parkinson's	33.2	Glob Palladus Alzheimer's2	20.2
BA7 Parkinson's2	61.1	Glob Palladus Parkinson's	76.8
BA7 Huntington's	51.4	Glob Palladus Parkinson's2	31.6
BA7 Huntington's2	69.3	Glob Palladus PSP	4.4
BA7 PSP	55.5	Glob Palladus PSP2	7.8
BA7 PSP2	39.8	Glob Palladus Depression	19.6
BA7 Depression	15.7	Temp Pole Control	25.7
BA9 Control	27.4	Temp Pole Control2	15.5
BA9 Control2	28.5	Temp Pole Alzheimer's	11.7
BA9 Alzheimer's	8.8	Temp Pole Alzheimer's2	16.4
BA9 Alzheimer's2	42.3	Temp Pole Parkinson's	34.9
BA9 Parkinson's	86.5	Temp Pole Parkinson's2	38.7
BA9 Parkinson's2	49.7	Temp Pole Huntington's	65.5
BA9 Huntington's	45.1	Temp Pole PSP	3.1
BA9 Huntington's2	40.1	Temp Pole PSP2	12.2
BA9 PSP	20.4	Temp Pole Depression2	19.1
BA9 PSP2	0.0	Cing Gyr Control	52.1
BA9 Depression	14.9	Cing Gyr Control2	14.1
BA9 Depression2	17.7	Cing Gyr Alzheimer's	13.1
BA17 Control	64.2	Cing Gyr Alzheimer's2	60.3

BA17 Control2	22.5	Cing Gyr Parkinson's	66.4
BA17 Alzheimer's2	52.1	Cing Gyr Parkinson's2	25.5
BA17 Parkinson's	85.3	Cing Gyr Huntington's	85.3
BA17 Parkinson's2	94.0	Cing Gyr Huntington's2	43.2
BA17 Huntington's	70.7	Cing Gyr PSP	26.2
BA17 Huntington's2	34.4	Cing Gyr PSP2	6.2
BA17 Depression	46.7	Cing Gyr Depression	12.9
BA17 Depression2	82.9	Cing Gyr Depression2	23.8

CNS_neurodegeneration_v1.0 Summary: Ag2987 The NOV43 gene exhibits significantly higher expression in the brains of Alzheimer's disease patients than normal controls. This is consistent with reports of increased purinoceptor expression in AD. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

References:

Moore D, Iritani S, Chambers J, Emson P. Immunohistochemical localization of the P2Y1 purinergic receptor in Alzheimer's disease. *Neuroreport* 2000 Nov 27;11(17):3799-803

The biological actions of extracellular nucleotides are mediated by two distinct classes of P2 receptor, P2X and P2Y. The G protein-coupled P2Y receptors comprise five mammalian subtypes, P2Y(1-11). The P2Y1 subtype is expressed abundantly throughout the human brain and is specifically localized to neuronal structures. In the present study, the distribution of the P2Y1 receptor was investigated in Alzheimer's disease (AD) brains. In contrast to control human brain, the P2Y1 receptor was localized to a number of characteristic AD structures such as neurofibrillary tangles, neuritic plaques and neuropil threads. Immunoblot analysis showed that this specific immunostaining observed over tangles was not a result of cross-reactivity between the anti-P2Y1 antiserum and abnormal tau protein, the major constituent of tangles. The significance of this altered P2Y1 cellular distribution in AD brains is at present unclear.

Panel 1.3D Summary: Ag2987 The NOV43 gene, a purinoceptor homolog, exhibits highly brain preferential expression in this panel. Purinoceptors found in GDNF sensitive sensory neurons mediate nociceptor function. Therefore, agents that block the action of this receptor may have utility in treating pain, either as analgesics or in inhibiting the establishment

of chronic pain. In addition, adenosine plays a significant neuromodulatory role in brain regions such as the hippocampus, cortex, basal ganglia, and thalamus. Thus, this purinoceptor is localized in a position to participate with the action of adenosine in these brain regions. The NOV43 gene product may also influence Ca^{2+} mobilization, a function performed by other purinoceptors. Ca^{2+} mobilization is an important component of the molecular process leading to neurotransmitter release, such as dopamine and glutamate. P2Y receptors have been shown to affect the release of dopamine, a critical neurotransmitter deficient in Parkinson's disease. P2 receptor agonists are known to induce secretion. Therefore, agents that modulate NOV43 may be effective treatments for Parkinson's disease via effecting enhanced dopamine release. Furthermore, glutamate is the main excitatory amino acid neurotransmitter. Glutamate exerts excitotoxic neuronal damage and death in a number of pathological conditions, including stroke. Therefore, agents that inhibit this gene product are likely to affect glutamate release in the brain and the subsequent cytotoxic action of glutamate in these regions. The overexpression of this gene in the brains of Alzheimer's disease patients in the CNS_neurodegeneration_v1.0 panel indicates that antagonists of this receptor may also have utility in countering the processes associated with this disease.

References:

Liu DM, Katnik C, Stafford M, Adams DJ. P2Y purinoceptor activation mobilizes intracellular Ca^{2+} and induces a membrane current in rat intracardiac neurones. *J Physiol* 2000 Jul 15;526 Pt 2:287-98

1. The mobilization of Ca^{2+} by purinoceptor activation and the relative contributions of intra- and extracellular sources of Ca^{2+} were investigated using microfluorimetric measurements of fura-2 loaded in cultured neurones from rat intracardiac ganglia.
2. Reverse transcriptase-polymerase chain reaction (RT-PCR) revealed expression of mRNA for the G protein-coupled P2Y2 and P2Y4 receptors.
3. Brief application of either 300 μM ATP or 300 μM UTP caused transient increases in $[\text{Ca}^{2+}]_i$ of 277 ± 22 nM and 267 ± 39 nM, respectively. Removal of external Ca^{2+} did not significantly reduce these $[\text{Ca}^{2+}]_i$ responses.
4. The order of purinoceptor agonist potency for $[\text{Ca}^{2+}]_i$ increases was $\text{ATP} = \text{UTP} > 2\text{-MeSATP} > \text{ADP} \gg \text{adenosine}$, consistent with the profile for P2Y2 purinoceptors. ATP- and UTP-induced rises in $[\text{Ca}^{2+}]_i$ were completely and reversibly blocked by 10 μM PPADS (a P2 purinoceptor antagonist) and partially inhibited by 100 μM suramin (a relatively non-specific purinoceptor antagonist).
5. In the presence of the endoplasmic reticulum Ca^{2+} -ATPase inhibitor cyclopiazonic acid (10 μM) in Ca^{2+} -free media, the $[\text{Ca}^{2+}]_i$ responses evoked by ATP were progressively decreased and abolished.
6. ATP- and UTP-induced

[Ca²⁺]_i rises were insensitive to pertussis toxin, caffeine (5 mM) and ryanodine (10 µM) but were significantly reduced by U-73122, a phospholipase C (PLC) inhibitor. 7. In fura-2-loaded cells, perforated patch whole-cell recordings show that ATP and UTP evoked slow outward currents at -60 mV, concomitant with the rise in [Ca²⁺]_i, in approximately 30 % of rat intracardiac neurones. 8. In conclusion, these results suggest that in r intracardiac neurones, ATP binds to P2Y₂ purinoceptors to transiently raise [Ca²⁺]_i and activate an outward current. The signalling pathway appears to involve a PTX-insensitive G protein coupled to PLC generation of IP₃ which triggers the release of Ca²⁺ from a ryanodine-insensitive Ca²⁺ store(s).

Driessen B, Bultmann R, Jurna I, Baldauf J. Depression of C fiber-evoked activity by intrathecally administered reactive red 2 in rat thalamic neurons. *Brain Res* 1998 Jun 15;796(1-2):284-90

To investigate the possible role of spinal purinoceptors in nociception, the potent P₂-purinoceptor antagonist reactive red 2 was studied in rats under urethane anesthesia in which nociceptive activity was elicited by electrical stimulation of afferent C fibers in the sural nerve and recorded from single neurons in the ventrobasal complex of the thalamus. Intrathecal (i.t.) application of reactive red 2 (6-200 micrograms) caused a dose-dependent reduction of the evoked activity in thalamic neurons. The estimated ED₅₀ was 30 micrograms, and the maximum depression of nociceptive activity amounted to about 70% of the control activity at a dose of 100 micrograms. Morphine, administered i.t. at a maximally effective dose (80 micrograms), inhibited the evoked nociceptive activity by only up to 55% of the control activity. An i.t. co-injection of reactive red 2 (100 micrograms) and morphine (80 micrograms) caused a maximum reduction of the evoked thalamic activity by up to 85% of the control activity, thus, exceeding significantly the effect elicited by either drug alone. Similarly, i.t. co-injection of almost equipotent dosages of reactive red 2 (30 micrograms) and morphine (30 micrograms) caused a maximum reduction of the evoked activity by up to 72% of the control activity, which again exceeded significantly the effect of either drug alone. The results suggest that in rats reactive red 2 exerts antinociception by blockade of P₂-purinoceptors in the spinal cord and, hence, support the idea that ATP may play an important role in spinal transmission of nociceptive signals. An activation of the spinal opioid system does not seem to contribute to the effect of reactive red 2 but might act additive or even synergistically with its antinociceptive action.

Krugel U, Kittner H, Franke H, Illes P. Stimulation of P2 receptors in the ventral tegmental area enhances dopaminergic mechanisms in vivo. *Neuropharmacology* 2001 Jun;40(8):1084-93

It has been shown that endogenous adenosine 5'-triphosphate (ATP) as well as its
 5 exogenously applied structural analogue, 2-methylthio ATP (2-MeSATP), facilitate the release
 of dopamine from axon terminals in the rat nucleus accumbens (NAc) by activating ATP-
 sensitive P2 receptors. In the present study, reversed microdialysis of 2-MeSATP (10 microM,
 100 microM and 1 mM), or its microinjection (0.5, 5.0 and 50 pmol) into the ventral tegmental
 area (VTA), dose-dependently increased the local extracellular level of dopamine and the
 10 locomotion in the open field, respectively. These effects were abolished by the P2-receptor
 antagonist pyridoxalphosphate-6-azophenyl-2',4'-disulfonic acid (PPADS). When applied
 alone, the antagonist decreased the basal dopamine concentration, indicating that endogenous
 ATP controls the somatodendritic release of dopamine. Repeated microinjections of 2-
 MeSATP (5 pmol) once daily for 4 days led to a reproducible locomotor stimulation in the
 15 open field. Conditioned locomotion was induced by re-exposure to the novel environment on
 the seventh day. A challenge with amphetamine (1 mg/kg intraperitoneally) on the eighth day
 enhanced the locomotor activity in the 2-MeSATP-treated group in the sense of a cross-
 sensitisation, but failed to do so in the control group. Neurons in the VTA were heavily stained
 with antibodies developed against the P2Y(1) subtype of P2 receptors. Taken together, our
 20 data suggest that P2 receptors (probably of the P2Y(1) subtype) are involved in the initiation
 of somatodendritic dopamine release in the VTA and thereby may have a profound influence
 on sensitisation and reward-motivated behaviour.

Fernandez-Alvarez J, Hillaire-Buys D, Loubatieres-Mariani MM, Gomis R, Petit P. P2
 25 receptor agonists stimulate insulin release from human pancreatic islets. *Pancreas* 2001
 Jan;22(1):69-71

Although P2 receptors for adenosine 5'-triphosphate (ATP) and/or adenosine 5'-
 diphosphate (ADP) have been characterized in mammalian pancreatic beta cells, no evidence
 for an insulin-secreting effect of P2 receptor agonists has been reported as yet in humans. The
 present study aimed at investigating whether P2 receptor agonists could stimulate insulin
 30 release in human pancreatic islets obtained from brain-dead organ donors. Experiments were
 performed using different glucose concentrations and insulin was measured by
 radioimmunoassay. When the glucose concentration (8.3 mmol/L) was slightly stimulating for
 insulin release, alpha,beta-methylene ATP (200 micromol/L) and ADPbetaS (50 micromol/L)
 similarly amplified insulin secretion: both compounds induced a threefold increase in insulin

response. In the presence of a nonstimulating glucose concentration (3.0 mmol/L), only alpha,beta-methylene ATP could induce a significant 1.4-fold increase in insulin release, ADPbetaS being completely ineffective. These results give evidence that P2 receptor agonists are effective in stimulating insulin release in humans, the effect of the P2Y agonist being essentially glucose dependent

Panel 4D Summary: Ag 2987 The NOV43 transcript is expressed in lung fibroblasts after treatment with IFNg, IL-4, IL-9 other cytokines (CTs=32). This gene is also expressed in small airway epithelium treated with the inflammatory cytokines TNF-a and IL-1. This expression profile suggests a role for this transcript in lung inflammation. Low but detectable expression of this transcript is found also in dermal fibroblasts, primary CD4 T cells, EOL and antigen presenting cells.

This transcript encodes for a PY2receptor like molecule. Expression of this receptor has been reported in several cell types including eosinophils (Ref.1) and lung epithelium where it has been shown to mediate Cl(-) secretion via an increase in intracellular calcium concentration (ref. 2). Thus, the NOV43 gene product may influence Ca2+ mobilization, a funtion performed by other purinoceptors, and therefore lead to activation or secretion processes. As suggested by ref.3, the release of nucleotides by damaged cells in inflammation can lead to the activation of purinoreceptors and of other cells present in the inflamed tissues, including lung epithelium in asthma, COPD, emphysema and the skin in psoriasis or other skin inflammatory diseases. This release can also result in the activation of antigen presenting cells and T cells, which contribute to the perpetuation of the inflammatory process. Therefore, modulation of the expression or activity of the protein encoded by the NOV43 gene may prevent or reduce the inflammation process in all of these diseases and other autoimmune diseases, including as inflammatory bowel diseases.

References:

Idzko M, Dichmann S, Panther E, Ferrari D, Herouy Y, Virchow C Jr, Luttmann W, Di Virgilio F, Norgauer J.

Functional characterization of P2Y and P2X receptors in human eosinophils. J Cell Physiol 2001 Sep;188(3):329-36

Activation of purinoceptor by ATP induces in eosinophils various cell responses including calcium transients, actin polymerization, production of reactive oxygen metabolites, CD11b-expression, and chemotaxis. Here, the effect of ion channel-gated P2X and/or G protein-coupled P2Y receptor agonists ATP, ATPgammaS, alpha,beta-meATP, 2-MeSATP, BzATP, ADP, CTP, and UTP on the intracellular Ca(2+)-mobilization, actin polymerization,

production of reactive oxygen metabolites, CD11b expression and chemotaxis of human eosinophils were measured and the biological activity was analyzed. Although all tested nucleotides were able to induce all these cell responses, the biological activity of the analyzed nucleotides were distinct. Agonists of the G protein-coupled P2Y receptors such as 2-MeSATP, UTP, and ADP have a higher biological activity for production of reactive oxygen metabolites, actin polymerization and chemotaxis in comparison to the ion channel-gated P2X agonists alphabeta-meATP, BzATP, and CTP. In contrast, P2Y and P2X agonist showed similar potencies in respect to intracellular calcium transient and CD11b up-regulation. This conclusion was further supported by experiments with receptor iso-type antagonist KN62, EGTA or with the G(i) protein-inactivating pertussis toxin. These findings indicate participation of different purinoreceptors in the regulation of cell responses in eosinophils.

Laubinger W, Streubel G, Reiser G. Physiological evidence for a P2Y receptor responsive to diadenosine polyphosphates in human lung via Ca(2+) release studies in bronchial epithelial cells. *Biochem Pharmacol* 2001 Mar 1;61(5):623-9

P2Y(2) receptors that are activated by the extracellular nucleotides ATP or UTP mediate Cl(-) secretion via an increase in [Ca(2+)](i) (intracellular calcium concentration). Therefore, in the lung of patients suffering from cystic fibrosis, inhalation of aerosolized UTP offers a way to circumvent the defect in Cl(-) secretion by the cystic fibrosis transmembrane conductance regulator. A possible alternative for the relatively unstable UTP in inhalation therapy is the more resistant diadenosine tetraphosphate (Ap(4)A). In human and rat lung membranes, Ap(4)A binds to P2 receptor sites coupled to G proteins. Here, we showed that Ap(4)A caused an increase in [Ca(2+)](i) with an EC(50) of 17 microM in human bronchial epithelial cells (HBE1). The [Ca(2+)](i) rise evoked by ATP and UTP was completely, but that induced by Ap(4)A only partially, caused by release of Ca(2+) from internal stores. Moreover, the potency of Ap(4)A to mobilize Ca(2+) was lower than that of ATP and UTP (EC(50) 1.5 and 1.8 microM, respectively), and the maximal increase in [Ca(2+)](i) was considerably smaller than that after ATP or UTP. In accordance with our previous results providing evidence for a common binding site for various diadenosine polyphosphates in lung membranes, all Ap(n)A analogues tested (n = 3 to 6) caused a comparable [Ca(2+)](i) increase. Homologous or heterologous prestimulation largely diminished the increase in [Ca(2+)](i) found after a second pulse of either UTP or Ap(4)A. Although specific binding characteristics and functional responses of Ap(4)A on lung cells are in favor of a distinct receptor for Ap(4)A, the cross-talk between UTP and Ap(4)A in HBE1 cells and the only slight differences in Ca(2+) mobilization by ATP or UTP and Ap(4)A render it impossible at

this point to state unequivocally whether there exists a distinct P2Y receptor specific for diadenosine polyphosphates in lung epithelia or whether Ap(4)A activates one of the nucleotide receptors already described.

Di Virgilio F, Chiozzi P, Ferrari D, Falzoni S, Sanz JM, Morelli A, Torboli M, Bolognesi G, Baricordi OR. Nucleotide receptors: an emerging family of regulatory molecules in blood cells. Blood 2001 Feb 1;97(3):587-600

Nucleotides are emerging as an ubiquitous family of extracellular signaling molecules. It has been known for many years that adenosine diphosphate is a potent platelet aggregating factor, but it is now clear that virtually every circulating cell is responsive to nucleotides. Effects as different as proliferation or differentiation, chemotaxis, release of cytokines or lysosomal constituents, and generation of reactive oxygen or nitrogen species are elicited upon stimulation of blood cells with extracellular adenosine triphosphate (ATP). These effects are mediated through a specific class of plasma membrane receptors called purinergic P2 receptors that, according to the molecular structure, are further subdivided into 2 subfamilies: P2Y and P2X. ATP and possibly other nucleotides are released from damaged cells or secreted via nonlytic mechanisms. Thus, during inflammation or vascular damage, nucleotides may provide an important mechanism involved in the activation of leukocytes and platelets. However, the cell physiology of these receptors is still at its dawn, and the precise function of the multiple P2X and P2Y receptor subtypes remains to be understood.

Panel CNS_1 Summary: Ag2987 The expression in this panel confirms expression of the NOV43 gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

NOV44

Expression of gene NOV44 was assessed using the primer-probe sets Ag2988 and Ag2989, described in Tables AMA and AMB. Results of the RTQ-PCR runs are shown in Table AMC.

Table AMA. Probe Name Ag2988

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctggccaacctatcctttattg-3'	22	195	1153
Probe	TET-5'-tggtcctctaaactcattgctgactca-3'-TAMRA	26	238	1154
Reverse	5'-agatgggtctctccctcatacaa-3'	22	264	1155

Table AMB. Probe Name Ag2989

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctggccaacctatcctttattg-3'	22	195	1156
Probe	TET-5'-tggtcctctaaactcattgctgactca-3'-TAMRA	26	238	1157
Reverse	5'-agatgggttctcccctcatacaa-3'	22	264	1158

Table AMC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2988, Run 164523397	Tissue Name	Rel. Exp.(%) Ag2988, Run 164523397
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0

LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	23.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	5.3	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti- CD40	0.0	IBD Colitis 2	19.9
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	10.2
HUVEC none	0.0	Kidney	0.0
HUVEC starved	24.5		

CNS_neurodegeneration_v1.0 Summary: Ag2988/Ag2989 Expression of the NOV44 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 1.3D Summary: Ag2988/Ag2989 Expression of the NOV44 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2988 Significant expression of this gene is detected in a liver cirrhosis sample (CT = 32.7). This gene encodes a putative GPCR; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative GPCR could also be used for the diagnosis of liver cirrhosis.

NOV45

Expression of gene NOV45 was assessed using the primer-probe sets Ag2979, Ag2982, Ag2981 and Ag2984, described in Tables ANA, ANB, ANC and AND. Results of the RTQ-PCR runs are shown in Table ANE.

Table ANA. Probe Name Ag2979

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tacttcttcgtgggctcctt-3'	20	827	1159
Probe	TET-5'-aaggcagaacctgaagctggttctcc-3'-TAMRA	26	862	1160
Reverse	5'-cattcacctcagtcgtgtcc-3'	20	901	1161

Table ANB. Probe Name Ag2982

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tacttcttcgtgggctcctt-3'	20	827	1162
Probe	TET-5'-aaggcagaacctgaagctggttctcc-3'-TAMRA	26	862	1163
Reverse	5'-cattcacctcagtcgtgtcc-3'	20	901	1164

Table ANC. Probe Name Ag2981

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tacttcttcgtgggctcctt-3'	20	827	1165
Probe	TET-5'-aaggcagaacctgaagctggttctcc-3'-TAMRA	26	862	1166
Reverse	5'-cattcacctcagtcgtgtcc-3'	20	901	1167

Table AND. Probe Name Ag2984

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atcctccatcccatcttcaa-3'	20	284	1168

Probe	TET-5'-cctcagccctgtgatgatgttttct-3'- TAMRA	26	307	1169
Reverse	5'-cgcttagaaagctcaggctt-3'	20	340	1170

Table ANE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2982, Run 158603041	Tissue Name	Rel. Exp.(%) Ag2982, Run 158603041
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.3	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	19.8
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	5.5
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	58.2
Primary Th2 rest	0.0	Small airway epithelium none	3.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	3.6	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.8
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	2.4	CCD1106 (Keratinocytes) none	0.3
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.8
LAK cells IL-2	0.0	Liver cirrhosis	23.3
LAK cells IL-2+IL-12	0.0	Lupus kidney	2.7

LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	4.8	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	5.7	HPAEC TNF alpha + IL-1 beta	0.7
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	1.8
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	3.1	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.7	Dermal fibroblast CCD1070 TNF alpha	4.1
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.4
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag2979/Ag2982 Expression of the NOV45 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 1.3D Summary: Ag2981/Ag2984 Expression of the NOV45 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2982 Expression of the NOV45 gene is restricted to a few samples, with highest expression in small airway epithelium treated with TNF-alpha and IL-1

beta (CT=31.2). Significant expression is treated in bronchial epithelium and lung microvascular endothelial cells. Thus, expression of this gene could be used as a marker for activated epithelium. The expression in lung derived samples suggests that this protein may be involved in lung inflammatory disorders, including asthma and chronic obstructive pulmonary disorder. Results from a second experiment with the probe/primer set Ag2979 are not included because the amp plot indicates that there is a potential problem in one of the sample wells.

NOV46a and NOV46b

Expression of gene NOV46s and variant NOV46b was assessed using the primer-probe sets Ag2990 and Ag2991, described in Tables AOA and AOB. Results of the RTQ-PCR runs are shown in Tables AOC. Please note that variant NOV46B does not match the probe and primer set Ag2990.

Table AOA. Probe Name Ag2990

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cggaactgaggagactctttg-3'	21	59	1171
Probe	TET-5'-tacaagcagaccttgagcctcacggt-3'-TAMRA	26	81	1172
Reverse	5'-gagcacaactgcgttttcct-3'	19	140	1173

Table AOB. Probe Name Ag2991

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tggaacaggggaagtcttattttg-3'	22	742	1174
Probe	TET-5'-tttctctgtcgcctcttaacagcagtg-3'-TAMRA	26	785	1175
Reverse	5'-agccacgaagaagtaaagtat-3'	22	819	1176

Table AOC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2990, Run 164524407	Rel. Exp.(%) Ag2991, Run 164315033	Tissue Name	Rel. Exp.(%) Ag2990, Run 164524407	Rel. Exp.(%) Ag2991, Run 164315033
Secondary Th1 act	0.0	0.0	HUVEC IL-1beta	0.0	0.0
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	0.0	0.0
Secondary Tr1 act	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0

Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	0.0	0.0
Secondary Th2 rest	0.0	0.0	HUVEC IL-11	0.0	0.0
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	0.0	0.0
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	3.1	23.2
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	0.0	0.0
Primary Tr1 act	0.0	0.0	Microsvasular Dermal EC TNFalpha + IL- 1beta	0.0	11.6
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	12.3	100.0
Primary Th2 rest	0.0	0.0	Small airway epithelium none	0.0	12.1
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL- 1beta	11.7	98.6
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	0.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	0.0	0.0
Secondary CD8 lymphocyte rest	0.0	0.0	Astrocytes TNFalpha + IL- 1beta	0.0	0.0
Secondary CD8 lymphocyte act	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	23.8	5.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	0.0	CCD1106 (Keratinocytes) none	0.0	0.0
LAK cells rest	0.0	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-	0.0	10.7

			Ibeta		
LAK cells IL-2	0.0	0.0	Liver cirrhosis	100.0	18.0
LAK cells IL-2+IL-12	0.0	0.0	Lupus kidney	0.0	0.0
LAK cells IL-2+IFN gamma	0.0	0.0	NCI-H292 none	0.0	0.0
LAK cells IL-2+IL-18	0.0	0.0	NCI-H292 IL-4	0.0	3.8
LAK cells PMA/ionomycin	0.0	0.0	NCI-H292 IL-9	0.0	0.0
NK Cells IL-2 rest	27.5	0.0	NCI-H292 IL-13	0.0	0.0
Two Way MLR 3 day	0.0	0.0	NCI-H292 IFN gamma	0.0	0.0
Two Way MLR 5 day	2.1	0.0	HPAEC none	0.0	0.0
Two Way MLR 7 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast none	0.0	0.0
PBMC PWM	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	2.0
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	0.0	0.0
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	0.0	0.0
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IL-13	0.0	0.0
B lymphocytes PWM	0.0	0.0	Lung fibroblast IFN gamma	0.0	0.0
B lymphocytes CD40L and IL-4	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0	4.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0	0.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	0.0	0.0
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	7.0	0.0
Dendritic cells anti-CD40	0.0	0.0	IBD Colitis 2	0.0	6.2
Monocytes rest	0.0	0.0	IBD Crohn's	0.0	0.0
Monocytes LPS	0.0	0.0	Colon	0.0	0.0

Macrophages rest	0.0	0.0	Lung	0.0	0.0
Macrophages LPS	0.0	4.8	Thymus	0.0	0.0
HUVEC none	0.0	0.0	Kidney	0.0	0.0
HUVEC starved	0.0	0.0			

CNS_neurodegeneration_v1.0 Summary: Ag2990/Ag2991 Expression of the NOV46a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 1.3D Summary: Ag2990/Ag2991 Expression of the NOV46a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

Panel 4D Summary: Ag2990/Ag2991 Expression of the NOV46a gene is restricted to liver cirrhosis and TNFalpha + IL1beta treated bronchial and small airway epithelium. This expression profile suggests that antibodies or small molecule therapeutics designed with the putative protein encoded by this gene could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this putative protein product could also be used for the diagnosis of liver cirrhosis. In addition, the expression of this gene in tissues derived from the lung suggests that this gene product may be involved in pathological and inflammatory lung disorders that include chronic obstructive pulmonary disease, asthma, allergy and emphysema. A second experiment with Ag290 shows low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV46d

Expression of gene NOV46d was assessed using the primer-probe sets Ag2992 and Ag513, described in Tables APA and APB. Results of the RTQ-PCR runs are shown in Tables APC and APD.

Table APA. Probe Name Ag2992

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agggctggtcttctcttct-3'	20	703	1177
Probe	TET-5'-ccctcagcattcagggattcctatt-3'- TAMRA	26	731	1178
Reverse	5'-agtcattccaaatccttctcgat-3'	22	764	1179

Table APB. Probe Name Ag513

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tatctgtggttctctgtgggttca-3'	24	600	1180
Probe	TET-5'- atccacatatgatcctgacaagcaggaccag-3'-	31	626	1181

	TAMRA			
Reverse	5'-tggtcagcgcatcttctg-3'	19	659	1182

Table APC. Panel 1.1

Tissue Name	Rel. Exp.(%) Ag513, Run 124882567	Tissue Name	Rel. Exp.(%) Ag513, Run 124882567
Adrenal gland	0.0	Renal ca. UO-31	0.0
Bladder	0.0	Renal ca. RXF 393	0.0
Brain (amygdala)	0.0	Liver	0.0
Brain (cerebellum)	0.0	Liver (fetal)	0.0
Brain (hippocampus)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (substantia nigra)	0.0	Lung	0.0
Brain (thalamus)	0.0	Lung (fetal)	0.0
Cerebral Cortex	0.0	Lung ca. (non-s.cell) HOP-62	0.0
Brain (fetal)	0.0	Lung ca. (large cell)NCI-H460	0.0
Brain (whole)	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (non-sm. cell) A549	0.0
astrocytoma SNB-75	0.0	Lung ca. (s.cell var.) SHP-77	0.0
astrocytoma SW1783	0.0	Lung ca. (small cell) LX-1	0.0
glioma U251	0.0	Lung ca. (small cell) NCI-H69	81.8
glioma SF-295	0.0	Lung ca. (squam.) SW 900	0.0
glioma SNB-19	0.0	Lung ca. (squam.) NCI-H596	1.4
glio/astro U87-MG	0.0	Lymph node	0.0
neuro*; met SK-N-AS	0.0	Spleen	0.0
Mammary gland	0.0	Thymus	0.0
Breast ca. BT-549	0.0	Ovary	0.0
Breast ca. MDA-N	0.0	Ovarian ca. IGROV- 1	0.0
Breast ca.* (pl.ef) T47D	0.0	Ovarian ca. OVCAR- 3	0.0
Breast ca.* (pl.ef)	0.0	Ovarian ca. OVCAR-	0.0

MCF-7		4	
Breast ca.* (pl.ef) MDA-MB-231	0.0	Ovarian ca. OVCAR-5	100.0
Small intestine	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Colon ca. HT29	0.0	Pancreas	0.0
Colon ca. CaCo-2	0.0	Pancreatic ca. CAPAN 2	0.0
Colon ca. HCT-15	1.5	Pituitary gland	0.0
Colon ca. HCT-116	0.0	Placenta	0.0
Colon ca. HCC-2998	0.0	Prostate	0.0
Colon ca. SW480	0.0	Prostate ca.* (bone met) PC-3	0.0
Colon ca.* SW620 (SW480 met)	0.0	Salivary gland	0.0
Stomach	0.0	Trachea	0.0
Gastric ca. (liver met) NCI-N87	0.0	Spinal cord	0.0
Heart	0.0	Testis	0.0
Skeletal muscle (Fetal)	0.0	Thyroid	0.0
Skeletal muscle	0.0	Uterus	0.0
Endothelial cells	0.0	Melanoma M14	0.0
Heart (Fetal)	0.0	Melanoma LOX IMVI	0.0
Kidney	0.0	Melanoma UACC-62	0.0
Kidney (fetal)	0.0	Melanoma SK-MEL-28	0.0
Renal ca. 786-0	0.0	Melanoma* (met) SK-MEL-5	0.0
Renal ca. A498	0.0	Melanoma Hs688(A).T	0.0
Renal ca. ACHN	0.0	Melanoma* (met) Hs688(B).T	0.0
Renal ca. TK-10	0.0		

Table APD. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag513, Run 129119406	Tissue Name	Rel. Exp.(%) Ag513, Run 129119406
Endothelial cells	0.0	Renal ca. 786-0	0.0
Heart (Fetal)	0.0	Renal ca. A498	0.0
Pancreas	0.0	Renal ca. RXF 393	0.0

Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal Gland	0.0	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.0
Salivary gland	0.0	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	100.0
Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	0.0	Lung ca. (large cell) NCI-H460	0.0
Spinal cord	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U87-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) HOP-62	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) NCI-H596	40.6
astrocytoma SNB-75	0.0	Mammary gland	0.0
glioma SNB-19	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
glioma SF-295	0.0	Breast ca.* (pl. ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal Muscle	0.0	Breast ca. MDA-N	0.0
Bone marrow	0.0	Ovary	0.0
Thymus	0.0	Ovarian ca. OVCAR-3	0.0
Spleen	0.0	Ovarian ca. OVCAR-4	0.0
Lymph node	0.0	Ovarian ca. OVCAR-	91.4

of the expression or function of this gene may be effective in the treatment of lung, ovarian, and colon cancers.

Panel 1.3D Summary: Ag2992 Expression of the NOV46d gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

5 **Panel 4D Summary:** Ag2992 Expression of the NOV46d gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV46c

Expression of gene NOV46c was assessed using the primer-probe set Ag2985, described in Table AQA.

10

Table AQA. Probe Name Ag2985

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tctgtcatcccatctccaaa-3'	20	285	1183
Probe	TET-5'-atcctcattcctgtgatgacctttct-3'- TAMRA	26	305	1184
Reverse	5'-tcattggcactcagaaagctc-3'	20	346	1185

CNS_neurodegeneration_v1.0 Summary: Ag2985 Expression of the NOV46c gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

15 **Panel 4D Summary:** Ag2985 Expression of the NOV46c gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV47

Expression of gene NOV47 was assessed using the primer-probe set Ag2993, described in Table ARA. Results of the RTQ-PCR runs are shown in Tables ARB, ARC and ARD.

20

Table ARA. Probe Name Ag2993

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaggggttactgctttcacagaa-3'	22	161	1186
Probe	TET-5'-tgacttcacatgccataatggcactg-3'- TAMRA	26	211	1187
Reverse	5'-ctccctgtagatggacttgct-3'	21	239	1188

Table ARB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2993,	Tissue Name	Rel. Exp.(%) Ag2993,
-------------	----------------------	-------------	----------------------

	Run 211009463		Run 211009463
AD 1 Hippo	11.9	Control (Path) 3 Temporal Ctx	10.8
AD 2 Hippo	43.2	Control (Path) 4 Temporal Ctx	55.5
AD 3 Hippo	5.3	AD 1 Occipital Ctx	38.4
AD 4 Hippo	15.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	7.1
AD 6 Hippo	33.7	AD 4 Occipital Ctx	33.2
Control 2 Hippo	25.0	AD 5 Occipital Ctx	21.6
Control 4 Hippo	18.4	AD 6 Occipital Ctx	16.0
Control (Path) 3 Hippo	5.3	Control 1 Occipital Ctx	2.2
AD 1 Temporal Ctx	2.9	Control 2 Occipital Ctx	39.2
AD 2 Temporal Ctx	42.3	Control 3 Occipital Ctx	16.3
AD 3 Temporal Ctx	11.9	Control 4 Occipital Ctx	6.4
AD 4 Temporal Ctx	45.4	Control (Path) 1 Occipital Ctx	64.6
AD 5 Inf Temporal Ctx	86.5	Control (Path) 2 Occipital Ctx	15.8
AD 5 Sup Temporal Ctx	39.5	Control (Path) 3 Occipital Ctx	3.1
AD 6 Inf Temporal Ctx	24.8	Control (Path) 4 Occipital Ctx	27.9
AD 6 Sup Temporal Ctx	30.1	Control 1 Parietal Ctx	15.1
Control 1 Temporal Ctx	8.7	Control 2 Parietal Ctx	31.9
Control 2 Temporal Ctx	26.6	Control 3 Parietal Ctx	26.6
Control 3 Temporal Ctx	13.5	Control (Path) 1 Parietal Ctx	13.3
Control 3 Temporal Ctx	6.2	Control (Path) 2 Parietal Ctx	45.7
Control (Path) 1 Temporal Ctx	38.4	Control (Path) 3 Parietal Ctx	6.7
Control (Path) 2 Temporal Ctx	34.9	Control (Path) 4 Parietal Ctx	36.3

Skeletal muscle (fetal)	3.0	Breast ca. MDA-N	13.4
Skeletal muscle	21.2	Ovary	3.8
Bone marrow	2.4	Ovarian ca. OVCAR-3	12.0
Thymus	22.4	Ovarian ca. OVCAR-4	39.5
Spleen	4.6	Ovarian ca. OVCAR-5	34.6
Lymph node	6.3	Ovarian ca. OVCAR-8	2.8
Colorectal	45.1	Ovarian ca. IGROV-1	12.4
Stomach	6.0	Ovarian ca.* (ascites) SK-OV-3	3.7
Small intestine	19.6	Uterus	0.0
Colon ca. SW480	60.3	Placenta	23.8
Colon ca.* SW620(SW480 met)	13.2	Prostate	5.4
Colon ca. HT29	6.0	Prostate ca.* (bone met)PC-3	24.0
Colon ca. HCT-116	13.5	Testis	11.8
Colon ca. CaCo-2	31.6	Melanoma Hs688(A).T	5.0
Colon ca. tissue(ODO3866)	41.2	Melanoma* (met) Hs688(B).T	7.6
Colon ca. HCC-2998	32.5	Melanoma UACC-62	25.7
Gastric ca.* (liver met) NCI-N87	11.3	Melanoma M14	12.8
Bladder	45.1	Melanoma LOX IMVI	6.7
Trachea	8.5	Melanoma* (met) SK-MEL-5	30.1
Kidney	3.7	Adipose	3.2

Table ARD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2993, Run 164315034	Tissue Name	Rel. Exp.(%) Ag2993, Run 164315034
Secondary Th1 act	8.4	HUVEC IL-1beta	3.8
Secondary Th2 act	11.8	HUVEC IFN gamma	4.5
Secondary Tr1 act	14.3	HUVEC TNF alpha + IFN gamma	4.2
Secondary Th1 rest	3.3	HUVEC TNF alpha + IL4	5.0
Secondary Th2 rest	6.0	HUVEC IL-11	5.7

Secondary Tr1 rest	3.8	Lung Microvascular EC none	3.8
Primary Th1 act	22.7	Lung Microvascular EC TNFalpha + IL-1beta	3.3
Primary Th2 act	16.0	Microvascular Dermal EC none	10.2
Primary Tr1 act	30.4	Microvascular Dermal EC TNFalpha + IL-1beta	5.5
Primary Th1 rest	39.0	Bronchial epithelium TNFalpha + IL1beta	18.0
Primary Th2 rest	31.9	Small airway epithelium none	5.2
Primary Tr1 rest	20.6	Small airway epithelium TNFalpha + IL-1beta	25.9
CD45RA CD4 lymphocyte act	10.0	Coronary artery SMC rest	4.5
CD45RO CD4 lymphocyte act	24.1	Coronary artery SMC TNFalpha + IL-1beta	1.5
CD8 lymphocyte act	17.4	Astrocytes rest	4.5
Secondary CD8 lymphocyte rest	13.0	Astrocytes TNFalpha + IL-1beta	5.0
Secondary CD8 lymphocyte act	10.7	KU-812 (Basophil) rest	14.0
CD4 lymphocyte none	9.0	KU-812 (Basophil) PMA/ionomycin	22.4
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.5	CCD1106 (Keratinocytes) none	8.7
LAK cells rest	13.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.1
LAK cells IL-2	14.4	Liver cirrhosis	6.0
LAK cells IL-2+IL-12	13.1	Lupus kidney	5.0
LAK cells IL-2+IFN gamma	15.8	NCI-H292 none	25.3
LAK cells IL-2+ IL-18	13.8	NCI-H292 IL-4	28.3
LAK cells PMA/ionomycin	3.9	NCI-H292 IL-9	31.2
NK Cells IL-2 rest	10.7	NCI-H292 IL-13	14.9
Two Way MLR 3 day	11.7	NCI-H292 IFN gamma	11.9
Two Way MLR 5 day	6.8	HPAEC none	4.9
Two Way MLR 7 day	6.0	HPAEC TNF alpha + IL-1 beta	4.7
PBMC rest	3.3	Lung fibroblast none	2.6
PBMC PWM	30.4	Lung fibroblast TNF alpha + IL-1 beta	1.4
PBMC PHA-L	20.9	Lung fibroblast IL-4	7.1

Ramos (B cell) none	22.1	Lung fibroblast IL-9	6.5
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	4.9
B lymphocytes PWM	51.4	Lung fibroblast IFN gamma	7.1
B lymphocytes CD40L and IL-4	24.8	Dermal fibroblast CCD1070 rest	9.7
EOL-1 dbcAMP	5.6	Dermal fibroblast CCD1070 TNF alpha	24.1
EOL-1 dbcAMP PMA/ionomycin	4.5	Dermal fibroblast CCD1070 IL-1 beta	3.9
Dendritic cells none	7.0	Dermal fibroblast IFN gamma	4.0
Dendritic cells LPS	5.7	Dermal fibroblast IL-4	5.9
Dendritic cells anti-CD40	8.7	IBD Colitis 2	0.2
Monocytes rest	3.0	IBD Crohn's	0.0
Monocytes LPS	1.6	Colon	8.2
Macrophages rest	11.3	Lung	7.3
Macrophages LPS	5.8	Thymus	27.0
HUVEC none	6.7	Kidney	23.0
HUVEC starved	15.6		

CNS_neurodegeneration_v1.0 Summary: Ag2993 This panel does not show differential expression of the NOV47 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag2993 Significant expression of the NOV47 gene is restricted to the brain, with expression in the thalamus (CT=34.4). This gene also shows strong expression in the brain in the previous panel suggesting that this gene product may be involved in the normal functioning of the brain. Thus, the protein encoded by this gene may represent a small molecule target for the treatment of neurologic diseases.

Panel 4D Summary: Ag2993 The NOV47 gene, a Peptidyl Prolyl Cis-Trans Isomerase A homolog, is a novel member of the family of receptors for the widely used immunosuppressants cyclosporin A and FK506 (see Wang et al., 2001). The NOV47 gene is expressed at moderate levels in many of the tissues in this panel and is expressed at a somewhat higher level (CT = 30.3) in ionomycin-stimulated Ramos B lymphocytes. Therefore, small molecule drugs that antagonize the activity of the NOV47 gene product may be useful as immunosuppressants to reduce or eliminate the symptoms in patients with

autoimmune or inflammatory conditions, such as Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

References:

- 5 Wang HC, Kim K, Bakhtiar R, Germanas JP. Structure-activity studies of ground- and transition-state analogue inhibitors of cyclophilin. J Med Chem 2001 Aug 2;44(16):2593-600

Peptidyl-prolyl isomerases (PPlases) are ubiquitous cellular enzymes that play roles in cellular signaling and protein folding. In addition, these proteins are the receptors for the widely used immunosuppressants cyclosporin A and FK506. We report the first structure-activity studies of de novo designed inhibitors of cyclophilin, the cellular target of cyclosporin A. Our mechanism-based inhibitors were modeled on the ground- and transition-state structures of proline-containing peptides, the natural substrates of the enzyme. Both ground-state analogues 1 and transition-state analogues 2 were prepared as single enantiomers from L-proline following a "self-reproduction of chirality" procedure. The binding affinities of the analogues for the active site of cyclophilin were measured by a fluorescence perturbation assay. While the transition-state analogues 2 did not display significant avidity for the active site ($K(d) = 77$ microM for 2b), several ground-state analogues bound to the enzyme with low micromolar affinity ($K(d) = 1.5$ microM for 1e). These results proclaim that properly designed small molecular weight molecules can form strong complexes with cyclophilin and may find use as probes in cell biology and as therapeutic agents.

NOV48a

Expression of gene NOV48a was assessed using the primer-probe set Ag3006, described in Table ASA. Results of the RTQ-PCR runs are shown in Tables ASB, ASC, ASD, ASE and ASF.

25

Table ASA. Probe Name Ag3006

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gggcttggaatagagaaacct-3'	21	2991	1189
Probe	TET-5'-caacttcctcaaagcccaaagccaag-3'-TAMRA	26	3037	1190
Reverse	5'-gaagccttgagccttgatttat-3'	22	3069	1191

Table ASB. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag3006, Run	Tissue Name	Rel. Exp.(%) Ag3006, Run
-------------	-----------------------------	-------------	-----------------------------

	211059882		211059882
110967 COPD-F	27.2	112427 Match Control Psoriasis-F	72.7
110980 COPD-F	18.8	112418 Psoriasis-M	20.2
110968 COPD-M	20.0	112723 Match Control Psoriasis-M	19.8
110977 COPD-M	71.7	112419 Psoriasis-M	32.8
110989 Emphysema-F	46.0	112424 Match Control Psoriasis-M	22.7
110992 Emphysema-F	18.3	112420 Psoriasis-M	74.2
110993 Emphysema-F	23.8	112425 Match Control Psoriasis-M	56.6
110994 Emphysema-F	20.7	104689 (MF) OA Bone-Backus	55.1
110995 Emphysema-F	19.9	104690 (MF) Adj "Normal" Bone-Backus	70.7
110996 Emphysema-F	4.3	104691 (MF) OA Synovium-Backus	42.0
110997 Asthma-M	6.4	104692 (BA) OA Cartilage-Backus	50.7
111001 Asthma-F	33.4	104694 (BA) OA Bone-Backus	32.5
111002 Asthma-F	33.0	104695 (BA) Adj "Normal" Bone-Backus	52.1
111003 Atopic Asthma-F	31.6	104696 (BA) OA Synovium-Backus	25.7
111004 Atopic Asthma-F	28.5	104700 (SS) OA Bone-Backus	10.4
111005 Atopic Asthma-F	11.1	104701 (SS) Adj "Normal" Bone-Backus	42.0
111006 Atopic Asthma-F	5.8	104702 (SS) OA Synovium-Backus	71.2
111417 Allergy-M	13.7	117093 OA Cartilage Rep7	23.8
112347 Allergy-M	6.2	112672 OA Bone5	57.8
112349 Normal Lung-F	3.8	112673 OA Synovium5	22.7
112357 Normal Lung-F	39.8	112674 OA Synovial Fluid cells5	27.5
112354 Normal Lung-M	22.1	117100 OA Cartilage Rep14	12.3
112374 Crohns-F	36.9	112756 OA Bone9	41.2
112389 Match Control Crohns-F	15.6	112757 OA Synovium9	100.0

112375 Crohns-F	20.6	112758 OA Synovial Fluid Cells9	19.2
112732 Match Control Crohns-F	17.9	117125 RA Cartilage Rep2	25.0
112725 Crohns-M	6.1	113492 Bone2 RA	19.8
112387 Match Control Crohns-M	14.9	113493 Synovium2 RA	7.7
112378 Crohns-M	4.8	113494 Syn Fluid Cells RA	14.5
112390 Match Control Crohns-M	48.0	113499 Cartilage4 RA	24.1
112726 Crohns-M	33.0	113500 Bone4 RA	25.9
112731 Match Control Crohns-M	25.2	113501 Synovium4 RA	15.4
112380 Ulcer Col-F	32.3	113502 Syn Fluid Cells4 RA	7.5
112734 Match Control Ulcer Col-F	32.5	113495 Cartilage3 RA	11.8
112384 Ulcer Col-F	35.6	113496 Bone3 RA	12.6
112737 Match Control Ulcer Col-F	6.5	113497 Synovium3 RA	5.6
112386 Ulcer Col-F	17.3	113498 Syn Fluid Cells3 RA	13.9
112738 Match Control Ulcer Col-F	11.3	117106 Normal Cartilage Rep20	14.2
112381 Ulcer Col-M	5.1	113663 Bone3 Normal	6.8
112735 Match Control Ulcer Col-M	41.2	113664 Synovium3 Normal	1.1
112382 Ulcer Col-M	25.2	113665 Syn Fluid Cells3 Normal	3.8
112394 Match Control Ulcer Col-M	11.2	117107 Normal Cartilage Rep22	15.0
112383 Ulcer Col-M	23.0	113667 Bone4 Normal	23.8
112736 Match Control Ulcer Col-M	13.0	113668 Synovium4 Normal	24.5
112423 Psoriasis-F	34.9	113669 Syn Fluid Cells4 Normal	37.9

Table ASC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3006, Run 165517770	Tissue Name	Rel. Exp.(%) Ag3006, Run 165517770
Liver adenocarcinoma	4.7	Kidney (fetal)	1.2
Pancreas	0.7	Renal ca. 786-0	0.8
Pancreatic ca. CAPAN	1.1	Renal ca. A498	2.5

2			
Adrenal gland	1.0	Renal ca. RXF 393	0.9
Thyroid	1.9	Renal ca. ACHN	2.2
Salivary gland	2.0	Renal ca. UO-31	1.7
Pituitary gland	7.0	Renal ca. TK-10	1.7
Brain (fetal)	2.2	Liver	0.2
Brain (whole)	4.0	Liver (fetal)	0.7
Brain (amygdala)	4.6	Liver ca. (hepatoblast) HepG2	1.8
Brain (cerebellum)	2.1	Lung	1.2
Brain (hippocampus)	3.6	Lung (fetal)	0.9
Brain (substantia nigra)	3.5	Lung ca. (small cell) LX-1	3.0
Brain (thalamus)	8.0	Lung ca. (small cell) NCI-H69	0.6
Cerebral Cortex	2.3	Lung ca. (s.cell var.) SHP-77	2.4
Spinal cord	5.2	Lung ca. (large cell)NCI-H460	6.1
glio/astro U87-MG	1.2	Lung ca. (non-sm. cell) A549	0.9
glio/astro U-118-MG	9.9	Lung ca. (non-s.cell) NCI-H23	2.8
astrocytoma SW1783	1.2	Lung ca. (non-s.cell) HOP-62	1.1
neuro*; met SK-N-AS	8.8	Lung ca. (non-s.cl) NCI-H522	2.1
astrocytoma SF-539	2.6	Lung ca. (squam.) SW 900	1.3
astrocytoma SNB-75	3.7	Lung ca. (squam.) NCI-H596	2.6
glioma SNB-19	2.7	Mammary gland	2.1
glioma U251	15.3	Breast ca.* (pl.ef) MCF-7	0.8
glioma SF-295	2.2	Breast ca.* (pl.ef) MDA-MB-231	3.1
Heart (fetal)	0.2	Breast ca.* (pl.ef) T47D	0.7
Heart	1.8	Breast ca. BT-549	2.7
Skeletal muscle (fetal)	4.4	Breast ca. MDA-N	1.6
Skeletal muscle	100.0	Ovary	1.2
Bone marrow	1.8	Ovarian ca. OVCAR- 3	1.7
Thymus	0.6	Ovarian ca. OVCAR- 4	2.2

Spleen	2.5	Ovarian ca. OVCAR-5	1.7
Lymph node	4.7	Ovarian ca. OVCAR-8	0.6
Colorectal	2.9	Ovarian ca. IGROV-1	0.9
Stomach	3.2	Ovarian ca.* (ascites) SK-OV-3	2.0
Small intestine	9.9	Uterus	3.2
Colon ca. SW480	0.9	Placenta	0.8
Colon ca.* SW620(SW480 met)	1.3	Prostate	4.2
Colon ca. HT29	0.2	Prostate ca.* (bone met)PC-3	1.6
Colon ca. HCT-116	2.0	Testis	6.3
Colon ca. CaCo-2	1.2	Melanoma Hs688(A).T	0.9
Colon ca. tissue(ODO3866)	0.9	Melanoma* (met) Hs688(B).T	0.3
Colon ca. HCC-2998	2.2	Melanoma UACC-62	3.1
Gastric ca.* (liver met) NCI-N87	3.9	Melanoma M14	5.0
Bladder	0.8	Melanoma LOX IMVI	0.7
Trachea	2.3	Melanoma* (met) SK-MEL-5	0.8
Kidney	4.0	Adipose	1.1

Table ASD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3006, Run 163577592	Tissue Name	Rel. Exp.(%) Ag3006, Run 163577592
Normal Colon	35.8	Kidney Margin 8120608	9.0
CC Well to Mod Diff (ODO3866)	5.6	Kidney Cancer 8120613	13.2
CC Margin (ODO3866)	6.2	Kidney Margin 8120614	16.2
CC Gr.2 rectosigmoid (ODO3868)	5.5	Kidney Cancer 9010320	8.7
CC Margin (ODO3868)	10.8	Kidney Margin 9010321	20.9
CC Mod Diff (ODO3920)	9.5	Normal Uterus	6.3
CC Margin (ODO3920)	19.6	Uterus Cancer 064011	9.4

CC Gr.2 ascend colon (ODO3921)	10.7	Normal Thyroid	21.0
CC Margin (ODO3921)	8.5	Thyroid Cancer 064010	2.5
CC from Partial Hepatectomy (ODO4309) Mets	6.1	Thyroid Cancer A302152	5.6
Liver Margin (ODO4309)	0.9	Thyroid Margin A302153	8.2
Colon mets to lung (OD04451-01)	6.8	Normal Breast	24.7
Lung Margin (OD04451-02)	2.8	Breast Cancer (OD04566)	65.1
Normal Prostate 6546-1	66.9	Breast Cancer (OD04590-01)	70.2
Prostate Cancer (OD04410)	16.6	Breast Cancer Mets (OD04590-03)	68.3
Prostate Margin (OD04410)	18.0	Breast Cancer Metastasis (OD04655-05)	100.0
Prostate Cancer (OD04720-01)	13.7	Breast Cancer 064006	14.7
Prostate Margin (OD04720-02)	25.5	Breast Cancer 1024	33.7
Normal Lung 061010	15.1	Breast Cancer 9100266	26.6
Lung Met to Muscle (ODO4286)	22.4	Breast Margin 9100265	15.3
Muscle Margin (ODO4286)	52.5	Breast Cancer A209073	16.7
Lung Malignant Cancer (OD03126)	9.9	Breast Margin A2090734	16.3
Lung Margin (OD03126)	11.2	Normal Liver	6.8
Lung Cancer (OD04404)	3.9	Liver Cancer 064003	1.6
Lung Margin (OD04404)	8.7	Liver Cancer 1025	2.1
Lung Cancer (OD04565)	6.5	Liver Cancer 1026	2.0
Lung Margin (OD04565)	6.2	Liver Cancer 6004-T	2.3
Lung Cancer (OD04237-01)	13.3	Liver Tissue 6004-N	3.2
Lung Margin (OD04237-02)	11.4	Liver Cancer 6005-T	2.6
Ocular Mel Met to Liver (ODO4310)	5.2	Liver Tissue 6005-N	0.6
Liver Margin (ODO4310)	2.0	Normal Bladder	9.7
Melanoma Mets to Lung (OD04321)	6.3	Bladder Cancer 1023	2.2

Lung Margin (OD04321)	11.5	Bladder Cancer A302173	4.4
Normal Kidney	47.6	Bladder Cancer (OD04718-01)	10.3
Kidney Ca, Nuclear grade 2 (OD04338)	27.7	Bladder Normal Adjacent (OD04718-03)	62.0
Kidney Margin (OD04338)	43.8	Normal Ovary	5.0
Kidney Ca Nuclear grade 1/2 (OD04339)	12.4	Ovarian Cancer 064008	17.9
Kidney Margin (OD04339)	46.7	Ovarian Cancer (OD04768-07)	20.3
Kidney Ca, Clear cell type (OD04340)	20.2	Ovary Margin (OD04768-08)	3.6
Kidney Margin (OD04340)	40.6	Normal Stomach	32.5
Kidney Ca, Nuclear grade 3 (OD04348)	4.0	Gastric Cancer 9060358	9.6
Kidney Margin (OD04348)	30.1	Stomach Margin 9060359	8.6
Kidney Cancer (OD04622-01)	9.1	Gastric Cancer 9060395	12.0
Kidney Margin (OD04622-03)	5.6	Stomach Margin 9060394	14.3
Kidney Cancer (OD04450-01)	9.1	Gastric Cancer 9060397	17.6
Kidney Margin (OD04450-03)	24.0	Stomach Margin 9060396	2.6
Kidney Cancer 8120607	18.3	Gastric Cancer 064005	20.7

Table ASE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3006, Run 170188143	Tissue Name	Rel. Exp.(%) Ag3006, Run 170188143
Daoy- Medulloblastoma	8.5	Ca Ski- Cervical epidermoid carcinoma (metastasis)	15.7
TE671- Medulloblastoma	7.5	ES-2- Ovarian clear cell carcinoma	8.3
D283 Med- Medulloblastoma	31.6	Ramos- Stimulated with PMA/ionomycin 6h	6.3
PFSK-1- Primitive Neuroectodermal	21.8	Ramos- Stimulated with PMA/ionomycin 14h	12.6
XF-498- CNS	5.4	MEG-01- Chronic	3.5

		myelogenous leukemia (megokaryoblast)	
SNB-78- Glioma	2.5	Raji- Burkitt's lymphoma	5.8
SF-268- Glioblastoma	13.4	Daudi- Burkitt's lymphoma	15.4
T98G- Glioblastoma	24.5	U266- B-cell plasmacytoma	11.4
SK-N-SH- Neuroblastoma (metastasis)	14.4	CA46- Burkitt's lymphoma	2.4
SF-295- Glioblastoma	8.0	RL- non-Hodgkin's B-cell lymphoma	1.3
Cerebellum	5.3	JM1- pre-B-cell lymphoma	6.4
Cerebellum	3.2	Jurkat- T cell leukemia	3.8
NCI-H292- Mucoepidermoid lung carcinoma	13.5	TF-1- Erythroleukemia	8.0
DMS-114- Small cell lung cancer	9.1	HUT 78- T-cell lymphoma	7.8
DMS-79- Small cell lung cancer	100.0	U937- Histiocytic lymphoma	8.3
NCI-H146- Small cell lung cancer	6.3	KU-812- Myelogenous leukemia	5.8
NCI-H526- Small cell lung cancer	20.6	769-P- Clear cell renal carcinoma	7.7
NCI-N417- Small cell lung cancer	21.5	Caki-2- Clear cell renal carcinoma	13.9
NCI-H82- Small cell lung cancer	15.2	SW 839- Clear cell renal carcinoma	0.9
NCI-H157- Squamous cell lung cancer (metastasis)	18.2	G401- Wilms' tumor	3.7
NCI-H1155- Large cell lung cancer	18.6	Hs766T- Pancreatic carcinoma (LN metastasis)	10.5
NCI-H1299- Large cell lung cancer	19.2	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	6.3
NCI-H727- Lung carcinoid	7.2	SU86.86- Pancreatic carcinoma (liver metastasis)	10.3
NCI-UMC-11- Lung carcinoid	12.5	BxPC-3- Pancreatic adenocarcinoma	3.7
LX-1- Small cell lung cancer	10.9	HPAC- Pancreatic adenocarcinoma	11.0
Colo-205- Colon cancer	7.1	MIA PaCa-2- Pancreatic carcinoma	1.4
KM12- Colon cancer	13.1	CFPAC-1- Pancreatic ductal adenocarcinoma	23.2
KM20L2- Colon cancer	1.1	PANC-1- Pancreatic	4.7

		epithelioid ductal carcinoma	
NCI-H716- Colon cancer	7.2	T24- Bladder carcinma (transitional cell)	3.5
SW-48- Colon adenocarcinoma	4.4	5637- Bladder carcinoma	4.7
SW1116- Colon adenocarcinoma	5.8	HT-1197- Bladder carcinoma	1.5
LS 174T- Colon adenocarcinoma	3.7	UM-UC-3- Bladder carcinma (transitional cell)	4.8
SW-948- Colon adenocarcinoma	0.5	A204- Rhabdomyosarcoma	35.4
SW-480- Colon adenocarcinoma	1.7	HT-1080- Fibrosarcoma	9.0
NCI-SNU-5- Gastric carcinoma	3.1	MG-63- Osteosarcoma	4.3
KATO III- Gastric carcinoma	24.8	SK-LMS-1- Leiomyosarcoma (vulva)	13.8
NCI-SNU-16- Gastric carcinoma	6.9	SJRH30- Rhabdomyosarcoma (met to bone marrow)	7.3
NCI-SNU-1- Gastric carcinoma	18.0	A431- Epidermoid carcinoma	2.8
RF-1- Gastric adenocarcinoma	4.1	WM266-4- Melanoma	5.4
RF-48- Gastric adenocarcinoma	3.4	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	10.0	MDA-MB-468- Breast adenocarcinoma	14.5
NCI-N87- Gastric carcinoma	6.1	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	3.5	SCC-15- Squamous cell carcinoma of tongue	0.3
HelaS3- Cervical adenocarcinoma	6.7	CAL 27- Squamous cell carcinoma of tongue	5.2

Table ASF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3006, Run 168033497	Tissue Name	Rel. Exp.(%) Ag3006, Run 168033497
Secondary Th1 act	32.3	HUVEC IL-1beta	9.0
Secondary Th2 act	25.5	HUVEC IFN gamma	21.9
Secondary Tr1 act	39.0	HUVEC TNF alpha + IFN gamma	29.9
Secondary Th1 rest	5.8	HUVEC TNF alpha + IL4	26.4

Secondary Th2 rest	7.8	HUVEC IL-11	11.0
Secondary Tr1 rest	15.9	Lung Microvascular EC none	27.7
Primary Th1 act	38.2	Lung Microvascular EC TNFalpha + IL-1beta	28.3
Primary Th2 act	33.7	Microvascular Dermal EC none	30.8
Primary Tr1 act	47.6	Microvascular Dermal EC TNFalpha + IL-1beta	17.4
Primary Th1 rest	35.8	Bronchial epithelium TNFalpha + IL1beta	23.3
Primary Th2 rest	25.7	Small airway epithelium none	12.3
Primary Tr1 rest	23.7	Small airway epithelium TNFalpha + IL-1beta	44.4
CD45RA CD4 lymphocyte act	23.7	Coronary artery SMC rest	13.2
CD45RO CD4 lymphocyte act	29.5	Coronary artery SMC TNFalpha + IL-1beta	10.4
CD8 lymphocyte act	33.7	Astrocytes rest	11.5
Secondary CD8 lymphocyte rest	37.6	Astrocytes TNFalpha + IL-1beta	6.0
Secondary CD8 lymphocyte act	16.4	KU-812 (Basophil) rest	14.5
CD4 lymphocyte none	4.8	KU-812 (Basophil) PMA/ionomycin	22.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	15.3	CCD1106 (Keratinocytes) none	28.1
LAK cells rest	11.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	14.8
LAK cells IL-2	29.9	Liver cirrhosis	2.2
LAK cells IL-2+IL-12	25.3	Lupus kidney	5.8
LAK cells IL-2+IFN gamma	44.1	NCI-H292 none	15.4
LAK cells IL-2+ IL-18	35.8	NCI-H292 IL-4	36.3
LAK cells PMA/ionomycin	8.7	NCI-H292 IL-9	22.8
NK Cells IL-2 rest	23.3	NCI-H292 IL-13	12.7
Two Way MLR 3 day	33.0	NCI-H292 IFN gamma	17.4
Two Way MLR 5 day	18.3	HPAEC none	22.8
Two Way MLR 7 day	17.0	HPAEC TNF alpha + IL-1 beta	15.9
PBMC rest	5.0	Lung fibroblast none	27.5
PBMC PWM	100.0	Lung fibroblast TNF alpha + IL-1 beta	14.3

PBMC PHA-L	31.2	Lung fibroblast IL-4	39.0
Ramos (B cell) none	30.4	Lung fibroblast IL-9	26.2
Ramos (B cell) ionomycin	66.4	Lung fibroblast IL-13	27.9
B lymphocytes PWM	96.6	Lung fibroblast IFN gamma	35.8
B lymphocytes CD40L and IL-4	32.3	Dermal fibroblast CCD1070 rest	31.2
EOL-1 dbcAMP	18.7	Dermal fibroblast CCD1070 TNF alpha	49.3
EOL-1 dbcAMP PMA/ionomycin	15.6	Dermal fibroblast CCD1070 IL-1 beta	9.7
Dendritic cells none	4.6	Dermal fibroblast IFN gamma	9.5
Dendritic cells LPS	3.0	Dermal fibroblast IL-4	15.7
Dendritic cells anti-CD40	4.0	IBD Colitis 2	6.0
Monocytes rest	7.5	IBD Crohn's	1.4
Monocytes LPS	10.8	Colon	25.0
Macrophages rest	6.8	Lung	7.9
Macrophages LPS	3.1	Thymus	55.1
HUVEC none	27.9	Kidney	18.4
HUVEC starved	35.8		

AI_comprehensive panel_v1.0 Summary: Ag3006 The NOV48a gene is a novel member of the Phospholipase C family that is expressed at moderate to low levels in numerous cell types involved in the immune response in health and disease. In addition, the NOV48a gene is expressed at higher levels (CTs range 29-32) in samples obtained from patients with several autoimmune and inflammatory diseases, particularly a subset of samples from osteoarthritic synovium and psoriasis. Therefore, small molecule drugs that antagonize the activity of this gene product may be useful as immunosuppressants to reduce or eliminate the symptoms in patients with conditions, such as osteoarthritis, psoriasis, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or lupus erythematosus.

Panel 1.3D Summary: Ag3006 Highest expression of the NOV48a gene, a phospholipase C homolog, is seen in skeletal muscle (CT=27.6). Phosphatidylinositol-specific phospholipase C (PLC) plays an important role in receptor-mediated signal transduction. In addition to skeletal muscle, this gene is expressed in heart, liver, adipose, adrenal, thyroid, and pancreas. This widespread expression in metabolic tissues suggests that this gene product may be involved in cellular regulation of metabolism through interactions with the insulin receptor.

A therapeutic modulator to this gene and/or gene-product may be useful in the treatment of metabolic diseases that affect skeletal muscle metabolism. This gene and/or gene-product may also prove useful in differentiating between fetal and adult forms of skeletal muscle tissue, since it is expressed at much higher levels in the adult (CT=27) when compared to expression in the fetal tissue (CT=32).

This gene is also expressed at a low level in almost all cancer cell lines in this panel. Hence, it is probably required for cell survival and proliferation and therefore, inhibition of this gene in cancer can probably be used as therapy.

In addition, increased expression of PLC delta has been observed in the brains of Alzheimer's disease patients, indicating a role for this class of enzyme in the disease process. Therefore, inhibitors of the NOV48A protein product, by countering this disease associated process, may have utility in treating Alzheimer's disease and other neurodegenerative disorders.

References:

Tanino H, Shimohama S, Sasaki Y, Sumida Y, Fujimoto S. Increase in phospholipase C-delta1 protein levels in aluminum-treated rat brains. *Biochem Biophys Res Commun* 2000 May 19;271(3):620-5

The effect of administration of aluminum to rats on the level of three phospholipase C (PLC) isozymes (beta1, gamma1, and delta1) was assessed in a variety of brain tissues. After exposure to aluminum, a statistically significant increase in malondialdehyde, an index of lipid peroxidation, was observed. In addition, there was a significant reduction in the catalytic activity of low molecular weight phosphotyrosine phosphatase, which loses its activity during oxidative stress. This suggests that oxidative stress is induced in brain tissues exposed to aluminum. The protein level of PLC-delta1, but not that of PLC-beta1 or -gamma1, was significantly increased in brains where oxidative stress had been induced. The total PLC activity in aluminum-treated rat brains was significantly higher than that in control brains. These results suggest that PLC-delta1 protein levels in brain tissues are increased by the induction of oxidative stress, giving an explanation for its up-regulation in Alzheimer's disease.

Panel 2D Summary: Ag3006 The NOV48A gene is expressed at a low to moderate level in most of the tissues on this panel. There is increased expression in ovarian and breast cancer compared to normal adjacent tissue. Thus, expression of this gene could potentially be used as a diagnostic marker for the presence of cancer. Furthermore, inhibition of this gene in ovarian and breast cancer may be useful as a therapeutic treatment. Additionally, there is

increased expression in normal kidney samples compared to adjacent tumors. Thus, decreased expression of this gene could be used as a diagnostic marker for kidney cancer and therapeutic modulation of expression of this gene in tumors may be used to treat these cancers.

Panel 3D Summary: Ag3006 The NOV48a gene is expressed at a low level in almost all cancer cell lines in this panel with the highest expression in DMS-79 (CT=29.21). This ubiquitous pattern of expression suggests that this gene product may be required for cell survival and proliferation and inhibition of this gene in cancer may therefore be useful as a therapy.

Panel 4D Summary: Ag3006 The NOV48a gene is ubiquitously expressed among the samples on this panel, suggesting a role for this protein product in inflammation. Please see AI_comprehensive panel_v1.0 for further discussion of utility of this gene in inflammation. Results from a second experiment with the CG56003-01 gene are not included. The amp plot indicates that there were experimental difficulties with this run.

NOV49

Expression of gene NOV49 was assessed using the primer-probe set Ag3003, described in Table ATA. Results of the RTQ-PCR runs are shown in Table ATB.

Table ATA. Probe Name Ag3003

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-catctcgtccaccctacgtt-3'	20	299	1192
Probe	TET-5'-cttcagctgctggtgcactcaaggat-3'-TAMRA	26	339	1193
Reverse	5'-ttcaggaagccatagaaactca-3'	22	366	1194

Table ATB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3003, Run 166245477	Tissue Name	Rel. Exp.(%) Ag3003, Run 166245477
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	15.2	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	17.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	14.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0

Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	17.7
Spinal cord	29.9	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	15.5
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	52.1
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	100.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.0
Thymus	0.0	Ovarian ca. OVCAR- 4	0.0
Spleen	0.0	Ovarian ca. OVCAR- 5	51.8
Lymph node	0.0	Ovarian ca. OVCAR- 8	0.0
Colorectal	0.0	Ovarian ca. IGROV- 1	0.0
Stomach	0.0	Ovarian ca.* (ascites)	0.0

		SK-OV-3	
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	11.2
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	31.2	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	13.3	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Panel 1.3D Summary: Ag3003 Expression of the NOV49 gene is restricted to a sample derived from a breast cancer cell line (CT=34.7). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of breast cancer. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of breast cancer.

Panel 4D Summary: Ag3003 Expression of the NOV49 gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.)

NOV50a

Expression of gene NOV50a was assessed using the primer-probe set Ag3014, described in Table AUA.

Table AUA. Probe Name Ag3014

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gactgagcgttgcccttttct-3'	20	1065	1195
Probe	TET-5'-agctacctcccaaagcagcctgacct-3'- TAMRA	26	1088	1196
Reverse	5'-acaatccctgcacaacgat-3'	19	1138	1197

CNS_neurodegeneration_v1.0 Summary: Ag3014 Expression of the NOV50a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 1.3D Summary: Ag3014 Expression of the NOV50a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

Panel 4D Summary: Ag3014 Expression of the NOV50a gene is low/undetectable in all samples on this panel (CTs>35). (Data not shown.) The amp plot indicates that there is a high probability of a probe failure.

10 NOV53

Expression of gene NOV53 was assessed using the primer-probe set Ag3008, described in Table AVA. Results of the RTQ-PCR runs are shown in Tables AVB, AVC and AVD.

Table AVA. Probe Name Ag3008

15

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cttaagctgctgcctatgaatg-3'	22	469	1198
Probe	TET-5'-atacgggagctacagaccatcatccg-3'-TAMRA	26	496	1199
Reverse	5'-tcacctctactggctgtcttgt-3'	22	524	1200

Table AVB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3008, Run 211010256	Tissue Name	Rel. Exp.(%) Ag3008, Run 211010256
AD 1 Hippo	14.7	Control (Path) 3 Temporal Ctx	6.2
AD 2 Hippo	30.6	Control (Path) 4 Temporal Ctx	35.4
AD 3 Hippo	6.3	AD 1 Occipital Ctx	22.1
AD 4 Hippo	8.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	84.1	AD 3 Occipital Ctx	4.9
AD 6 Hippo	57.0	AD 4 Occipital Ctx	22.4
Control 2 Hippo	28.9	AD 5 Occipital Ctx	72.2
Control 4 Hippo	22.1	AD 6 Occipital Ctx	21.5
Control (Path) 3 Hippo	4.7	Control 1 Occipital Ctx	4.7

AD 1 Temporal Ctx	26.1	Control 2 Occipital Ctx	64.6
AD 2 Temporal Ctx	30.8	Control 3 Occipital Ctx	10.7
AD 3 Temporal Ctx	5.9	Control 4 Occipital Ctx	12.2
AD 4 Temporal Ctx	17.6	Control (Path) 1 Occipital Ctx	73.7
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	10.5
AD 5 Sup Temporal Ctx	50.0	Control (Path) 3 Occipital Ctx	3.7
AD 6 Inf Temporal Ctx	57.0	Control (Path) 4 Occipital Ctx	9.7
AD 6 Sup Temporal Ctx	54.7	Control 1 Parietal Ctx	5.0
Control 1 Temporal Ctx	3.9	Control 2 Parietal Ctx	42.3
Control 2 Temporal Ctx	46.3	Control 3 Parietal Ctx	25.5
Control 3 Temporal Ctx	17.9	Control (Path) 1 Parietal Ctx	95.9
Control 3 Temporal Ctx	8.0	Control (Path) 2 Parietal Ctx	33.4
Control (Path) 1 Temporal Ctx	58.2	Control (Path) 3 Parietal Ctx	2.7
Control (Path) 2 Temporal Ctx	33.2	Control (Path) 4 Parietal Ctx	32.8

Table AVC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3008, Run 167927168	Tissue Name	Rel. Exp.(%) Ag3008, Run 167927168
Liver adenocarcinoma	18.9	Kidney (fetal)	42.6
Pancreas	7.6	Renal ca. 786-0	18.4
Pancreatic ca. CAPAN 2	9.4	Renal ca. A498	11.3
Adrenal gland	10.8	Renal ca. RXF 393	5.7
Thyroid	8.1	Renal ca. ACHN	5.2
Salivary gland	5.9	Renal ca. UO-31	9.1
Pituitary gland	11.6	Renal ca. TK-10	10.3
Brain (fetal)	31.9	Liver	6.2
Brain (whole)	30.1	Liver (fetal)	10.7
Brain (amygdala)	17.8	Liver ca. (hepatoblast) HepG2	17.3

Brain (cerebellum)	25.3	Lung	5.5
Brain (hippocampus)	25.7	Lung (fetal)	10.2
Brain (substantia nigra)	24.7	Lung ca. (small cell) LX-1	21.0
Brain (thalamus)	14.1	Lung ca. (small cell) NCI-H69	1.4
Cerebral Cortex	26.2	Lung ca. (s.cell var.) SHP-77	94.0
Spinal cord	10.5	Lung ca. (large cell)NCI-H460	4.7
glio/astro U87-MG	26.2	Lung ca. (non-sm. cell) A549	36.6
glio/astro U-118-MG	22.4	Lung ca. (non-s.cell) NCI-H23	12.7
astrocytoma SW1783	82.4	Lung ca. (non-s.cell) HOP-62	19.1
neuro*; met SK-N-AS	20.7	Lung ca. (non-s.cl) NCI-H522	33.2
astrocytoma SF-539	30.6	Lung ca. (squam.) SW 900	16.0
astrocytoma SNB-75	33.0	Lung ca. (squam.) NCI-H596	1.1
glioma SNB-19	22.7	Mammary gland	13.0
glioma U251	100.0	Breast ca.* (pl.ef) MCF-7	21.8
glioma SF-295	38.2	Breast ca.* (pl.ef) MDA-MB-231	38.7
Heart (fetal)	12.5	Breast ca.* (pl.ef) T47D	21.3
Heart	14.7	Breast ca. BT-549	15.6
Skeletal muscle (fetal)	10.4	Breast ca. MDA-N	21.6
Skeletal muscle	35.4	Ovary	15.3
Bone marrow	7.4	Ovarian ca. OVCAR- 3	11.6
Thymus	21.6	Ovarian ca. OVCAR- 4	10.2
Spleen	9.1	Ovarian ca. OVCAR- 5	12.4
Lymph node	16.8	Ovarian ca. OVCAR- 8	8.5
Colorectal	15.0	Ovarian ca. IGROV- 1	6.8
Stomach	17.8	Ovarian ca.* (ascites) SK-OV-3	58.2
Small intestine	4.7	Uterus	8.1

Colon ca. SW480	12.5	Placenta	1.3
Colon ca.* SW620(SW480 met)	85.9	Prostate	7.2
Colon ca. HT29	7.0	Prostate ca.* (bone met)PC-3	22.8
Colon ca. HCT-116	13.3	Testis	0.6
Colon ca. CaCo-2	32.8	Melanoma Hs688(A).T	10.1
Colon ca. tissue(ODO3866)	9.3	Melanoma* (met) Hs688(B).T	8.1
Colon ca. HCC-2998	19.9	Melanoma UACC-62	22.4
Gastric ca.* (liver met) NCI-N87	10.7	Melanoma M14	6.3
Bladder	21.6	Melanoma LOX IMVI	18.2
Trachea	9.3	Melanoma* (met) SK-MEL-5	12.5
Kidney	31.2	Adipose	23.0

Table AVD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3008, Run 164043360	Tissue Name	Rel. Exp.(%) Ag3008, Run 164043360
Secondary Th1 act	21.0	HUVEC IL-1beta	7.5
Secondary Th2 act	23.7	HUVEC IFN gamma	10.7
Secondary Tr1 act	23.5	HUVEC TNF alpha + IFN gamma	6.8
Secondary Th1 rest	8.2	HUVEC TNF alpha + IL4	7.5
Secondary Th2 rest	8.8	HUVEC IL-11	7.7
Secondary Tr1 rest	9.1	Lung Microvascular EC none	12.0
Primary Th1 act	24.8	Lung Microvascular EC TNFalpha + IL-1beta	8.2
Primary Th2 act	21.9	Microvascular Dermal EC none	17.9
Primary Tr1 act	29.1	Microvascular Dermal EC TNFalpha + IL-1beta	11.0
Primary Th1 rest	44.8	Bronchial epithelium TNFalpha + IL1beta	16.4
Primary Th2 rest	21.3	Small airway epithelium none	6.0
Primary Tr1 rest	20.4	Small airway epithelium TNFalpha + IL-1beta	48.3
CD45RA CD4	15.8	Coronary artery SMC rest	19.9

lymphocyte act			
CD45RO CD4 lymphocyte act	30.1	Coronary artery SMC TNFalpha + IL-1beta	8.0
CD8 lymphocyte act	22.8	Astrocytes rest	13.4
Secondary CD8 lymphocyte rest	32.5	Astrocytes TNFalpha + IL-1beta	8.9
Secondary CD8 lymphocyte act	15.7	KU-812 (Basophil) rest	9.5
CD4 lymphocyte none	8.8	KU-812 (Basophil) PMA/ionomycin	33.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	17.0	CCD1106 (Keratinocytes) none	12.3
LAK cells rest	21.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	8.6
LAK cells IL-2	19.5	Liver cirrhosis	2.1
LAK cells IL-2+IL-12	17.6	Lupus kidney	1.4
LAK cells IL-2+IFN gamma	28.9	NCI-H292 none	19.8
LAK cells IL-2+ IL-18	16.2	NCI-H292 IL-4	25.9
LAK cells PMA/ionomycin	23.2	NCI-H292 IL-9	27.0
NK Cells IL-2 rest	11.0	NCI-H292 IL-13	10.7
Two Way MLR 3 day	23.2	NCI-H292 IFN gamma	11.8
Two Way MLR 5 day	16.3	HPAEC none	6.0
Two Way MLR 7 day	9.6	HPAEC TNF alpha + IL-1 beta	9.0
PBMC rest	10.7	Lung fibroblast none	7.5
PBMC PWM	65.1	Lung fibroblast TNF alpha + IL-1 beta	6.7
PBMC PHA-L	58.6	Lung fibroblast IL-4	22.1
Ramos (B cell) none	33.0	Lung fibroblast IL-9	18.2
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	14.7
B lymphocytes PWM	74.7	Lung fibroblast IFN gamma	23.2
B lymphocytes CD40L and IL-4	23.2	Dermal fibroblast CCD1070 rest	18.8
EOL-1 dbcAMP	12.2	Dermal fibroblast CCD1070 TNF alpha	46.7
EOL-1 dbcAMP PMA/ionomycin	13.0	Dermal fibroblast CCD1070 IL-1 beta	8.0
Dendritic cells none	16.7	Dermal fibroblast IFN gamma	6.7
Dendritic cells LPS	13.0	Dermal fibroblast IL-4	12.2
Dendritic cells anti-	17.1	IBD Colitis 2	2.0

CD40			
Monocytes rest	17.2	IBD Crohn's	2.9
Monocytes LPS	7.5	Colon	15.4
Macrophages rest	24.5	Lung	14.8
Macrophages LPS	11.8	Thymus	22.2
HUVEC none	9.4	Kidney	19.8
HUVEC starved	18.8		

CNS_neurodegeneration_v1.0 Summary: Ag3008 This panel does not show differential expression of the NOV53 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3008 Highest expression of the NOV53 gene is seen in a brain cancer cell line (CT=29). In addition, this gene has low to moderate expression in all the cancer cell lines used in this panel. Thus, expression of this gene might be used as a diagnostic marker in brain, colon, renal, lung, melanoma and ovarian cancers.

This gene encodes a homolog of uracil phosphoribosyltransferase. This gene has low to moderate expression in several endocrine/metabolically-related tissues, including; adipose, adrenal, pancreas, liver and skeletal muscle. Therefore, a therapeutic modulator to this gene and/or gene-product may prove useful in the treatment of diseases which affect the endocrine system.

In addition, this gene shows moderate to low levels in the CNS and may be a small molecule target for the treatment of neurologic diseases.

Panel 4D Summary: Ag3008 The NOV53 gene, a uracil phosphoribosyl-transferase homolog is expressed at moderate to low levels in numerous cell types involved in the immune response. Higher levels of expression are seen in activated B lymphocytes, represented by ionomycin-activated Ramos (CT=27.6), and pokeweed mitogen-activated B lymphocytes (CT=28.02). Therefore, small molecules that antagonize the function of this gene product may be useful as therapeutic drugs to reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases in which B cells play a part in the initiation or progression of the disease process, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

NOV54a

Expression of gene NOV54a was assessed using the primer-probe sets Ag3015 and Ag3070, described in Tables AWA and AWB. Results of the RTQ-PCR runs are shown in Tables AWC, AWD, AWE and AWF.

Table AWA. Probe Name Ag3015

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtgctctcactatccacctcaa-3'	22	1515	1201
Probe	TET-5'-cacacatccatctcaagaggaacatt-3'- TAMRA	26	1537	1202
Reverse	5'-ccatacacttcagctctgact-3'	22	1573	1203

5

Table AWB. Probe Name Ag3070

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtgctctcactatccacctcaa-3'	22	1515	1204
Probe	TET-5'-cacacatccatctcaagaggaacatt-3'- TAMRA	26	1537	1205
Reverse	5'-ccatacacttcagctctgact-3'	22	1573	1206

Table AWC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3015, Run 211010356	Rel. Exp.(%) Ag3070, Run 208974108	Tissue Name	Rel. Exp.(%) Ag3015, Run 211010356	Rel. Exp.(%) Ag3070, Run 208974108
AD 1 Hippo	37.1	22.5	Control (Path) 3 Temporal Ctx	11.2	10.5
AD 2 Hippo	56.3	50.0	Control (Path) 4 Temporal Ctx	50.0	29.9
AD 3 Hippo	24.8	22.7	AD 1 Occipital Ctx	15.5	13.6
AD 4 Hippo	15.6	8.5	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	84.7	54.0	AD 3 Occipital Ctx	17.7	9.4
AD 6 Hippo	100.0	66.4	AD 4 Occipital Ctx	17.3	11.3
Control 2	37.4	32.5	AD 5	29.7	39.2

Hippo			Occipital Ctx		
Control 4 Hippo	33.0	26.1	AD 6 Occipital Ctx	45.4	13.1
Control (Path) 3 Hippo	15.2	8.8	Control 1 Occipital Ctx	9.7	8.2
AD 1 Temporal Ctx	31.2	17.4	Control 2 Occipital Ctx	42.0	30.4
AD 2 Temporal Ctx	50.3	38.2	Control 3 Occipital Ctx	15.5	14.5
AD 3 Temporal Ctx	17.6	14.8	Control 4 Occipital Ctx	13.8	9.4
AD 4 Temporal Ctx	36.1	23.3	Control (Path) 1 Occipital Ctx	92.7	67.8
AD 5 Inf Temporal Ctx	87.1	69.7	Control (Path) 2 Occipital Ctx	17.7	11.3
AD 5 Sup Temporal Ctx	87.1	53.2	Control (Path) 3 Occipital Ctx	7.9	8.9
AD 6 Inf Temporal Ctx	60.7	72.7	Control (Path) 4 Occipital Ctx	21.8	14.5
AD 6 Sup Temporal Ctx	55.9	35.8	Control 1 Parietal Ctx	16.5	10.4
Control 1 Temporal Ctx	15.9	8.0	Control 2 Parietal Ctx	77.9	48.6
Control 2 Temporal Ctx	56.6	28.9	Control 3 Parietal Ctx	20.3	19.3
Control 3 Temporal Ctx	28.3	17.8	Control (Path) 1 Parietal Ctx	76.3	100.0
Control 4 Temporal Ctx	21.5	16.7	Control (Path) 2 Parietal Ctx	36.1	21.2
Control (Path) 1 Temporal Ctx	82.4	70.2	Control (Path) 3 Parietal Ctx	7.1	9.3

Control (Path) 2 Temporal Ctx	49.7	38.7	Control (Path) 4 Parietal Ctx	41.8	30.6
----------------------------------	------	------	-------------------------------------	------	------

Table AWD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3015, Run 167927212	Rel. Exp.(%) Ag3070, Run 167985243	Tissue Name	Rel. Exp.(%) Ag3015, Run 167927212	Rel. Exp.(%) Ag3070, Run 167985243
Liver adenocarcinoma	23.7	21.3	Kidney (fetal)	56.3	64.6
Pancreas	7.4	7.2	Renal ca. 786- 0	2.3	2.9
Pancreatic ca. CAPAN 2	4.5	2.4	Renal ca. A498	13.9	11.7
Adrenal gland	6.5	4.9	Renal ca. RXF 393	14.1	25.7
Thyroid	14.6	14.6	Renal ca. ACHN	13.9	19.8
Salivary gland	13.3	1.4	Renal ca. UO- 31	7.4	17.8
Pituitary gland	6.0	8.4	Renal ca. TK- 10	12.6	14.0
Brain (fetal)	25.3	14.6	Liver	13.8	2.4
Brain (whole)	21.6	11.1	Liver (fetal)	5.7	4.9
Brain (amygdala)	28.1	32.3	Liver ca. (hepatoblast) HepG2	3.7	4.3
Brain (cerebellum)	17.9	17.2	Lung	23.7	7.6
Brain (hippocampus)	16.8	16.5	Lung (fetal)	20.6	33.0
Brain (substantia nigra)	18.9	30.6	Lung ca. (small cell) LX-1	8.8	15.2
Brain (thalamus)	9.1	2.8	Lung ca. (small cell) NCI-H69	3.5	10.4
Cerebral Cortex	42.6	36.9	Lung ca. (s.cell var.) SHP-77	13.1	17.3
Spinal cord	18.9	18.4	Lung ca. (large cell)NCI-H460	4.7	2.2
glio/astro U87-MG	19.8	37.6	Lung ca. (non- sm. cell) A549	13.1	10.2
glio/astro U-118- MG	11.9	18.3	Lung ca. (non- s.cell) NCI-	5.5	18.7

			H23		
astrocytoma SW1783	5.6	10.9	Lung ca. (non-s.cell) HOP-62	7.5	9.3
neuro*; met SK-N-AS	3.4	7.1	Lung ca. (non-s.cl) NCI-H522	18.8	34.6
astrocytoma SF-539	6.4	10.4	Lung ca. (squam.) SW 900	2.0	4.8
astrocytoma SNB-75	12.9	10.1	Lung ca. (squam.) NCI-H596	7.6	8.2
glioma SNB-19	13.3	16.0	Mammary gland	19.5	3.8
glioma U251	24.5	36.9	Breast ca.* (pl.ef) MCF-7	0.9	2.6
glioma SF-295	13.4	18.2	Breast ca.* (pl.ef) MDA-MB-231	8.0	9.7
Heart (fetal)	100.0	100.0	Breast ca.* (pl.ef) T47D	26.8	40.6
Heart	12.4	14.0	Breast ca. BT-549	5.1	6.1
Skeletal muscle (fetal)	51.1	67.8	Breast ca. MDA-N	11.0	16.2
Skeletal muscle	11.4	14.1	Ovary	50.0	53.2
Bone marrow	28.9	11.1	Ovarian ca. OVCAR-3	8.3	3.3
Thymus	53.2	56.3	Ovarian ca. OVCAR-4	11.0	18.7
Spleen	24.8	30.6	Ovarian ca. OVCAR-5	20.2	38.4
Lymph node	60.7	47.6	Ovarian ca. OVCAR-8	5.0	8.1
Colorectal	17.1	15.5	Ovarian ca. IGROV-1	4.7	7.9
Stomach	8.9	9.7	Ovarian ca.* (ascites) SK-OV-3	8.5	14.1
Small intestine	11.9	12.8	Uterus	25.5	26.2
Colon ca. SW480	17.9	16.4	Placenta	3.4	5.4
Colon ca.* SW620(SW480 met)	21.9	16.0	Prostate	14.3	13.5
Colon ca. HT29	2.4	2.3	Prostate ca.* (bone met)PC-	14.6	28.3

			3		
Colon ca. HCT-116	3.4	6.8	Testis	6.6	10.3
Colon ca. CaCo-2	11.7	16.0	Melanoma Hs688(A).T	12.9	12.4
Colon ca. tissue(ODO3866)	9.9	18.0	Melanoma* (met) Hs688(B).T	10.7	16.6
Colon ca. HCC-2998	15.0	17.8	Melanoma UACC-62	7.3	10.4
Gastric ca.* (liver met) NCI-N87	3.1	4.9	Melanoma M14	3.4	7.4
Bladder	10.5	5.8	Melanoma LOX IMVI	14.1	15.7
Trachea	13.2	10.9	Melanoma* (met) SK-MEL-5	2.5	2.8
Kidney	26.1	15.0	Adipose	34.2	33.0

Table AWE. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3070, Run 173800588	Tissue Name	Rel. Exp.(%) Ag3070, Run 173800588
Normal Colon	25.0	Kidney Margin (OD04348)	75.8
Colon cancer (OD06064)	74.2	Kidney malignant cancer (OD06204B)	5.0
Colon Margin (OD06064)	36.3	Kidney normal adjacent tissue (OD06204E)	11.9
Colon cancer (OD06159)	4.9	Kidney Cancer (OD04450-01)	44.4
Colon Margin (OD06159)	17.4	Kidney Margin (OD04450-03)	15.8
Colon cancer (OD06297-04)	1.4	Kidney Cancer 8120613	4.0
Colon Margin (OD06297-015)	18.9	Kidney Margin 8120614	37.4
CC Gr.2 ascend colon (ODO3921)	6.5	Kidney Cancer 9010320	19.3
CC Margin (ODO3921)	16.5	Kidney Margin 9010321	16.6
Colon cancer metastasis (OD06104)	12.7	Kidney Cancer 8120607	33.2
Lung Margin (OD06104)	34.6	Kidney Margin 8120608	18.4

Colon mets to lung (OD04451-01)	8.1	Normal Uterus	76.3
Lung Margin (OD04451-02)	52.5	Uterine Cancer 064011	12.8
Normal Prostate	19.1	Normal Thyroid	11.8
Prostate Cancer (OD04410)	0.0	Thyroid Cancer 064010	7.2
Prostate Margin (OD04410)	13.2	Thyroid Cancer A302152	37.4
Normal Ovary	60.7	Thyroid Margin A302153	9.7
Ovarian cancer (OD06283-03)	14.8	Normal Breast	100.0
Ovarian Margin (OD06283-07)	24.1	Breast Cancer (OD04566)	13.4
Ovarian Cancer 064008	26.6	Breast Cancer 1024	34.6
Ovarian cancer (OD06145)	25.2	Breast Cancer (OD04590-01)	13.3
Ovarian Margin (OD06145)	24.1	Breast Cancer Mets (OD04590-03)	40.6
Ovarian cancer (OD06455-03)	0.0	Breast Cancer Metastasis (OD04655- 05)	52.9
Ovarian Margin (OD06455-07)	22.4	Breast Cancer 064006	29.3
Normal Lung	36.9	Breast Cancer 9100266	28.7
Invasive poor diff. lung adeno (ODO4945-01)	19.1	Breast Margin 9100265	35.6
Lung Margin (ODO4945-03)	44.1	Breast Cancer A209073	10.7
Lung Malignant Cancer (OD03126)	18.0	Breast Margin A2090734	19.5
Lung Margin (OD03126)	18.3	Breast cancer (OD06083)	30.4
Lung Cancer (OD05014A)	22.5	Breast cancer node metastasis (OD06083)	59.5
Lung Margin (OD05014B)	79.6	Normal Liver	19.1
Lung cancer (OD06081)	9.5	Liver Cancer 1026	17.3
Lung Margin (OD06081)	17.7	Liver Cancer 1025	17.4
Lung Cancer (OD04237-01)	4.1	Liver Cancer 6004-T	27.0
Lung Margin (OD04237-02)	54.7	Liver Tissue 6004-N	8.7
Ocular Melanoma	14.3	Liver Cancer 6005-T	48.6

Metastasis			
Ocular Melanoma Margin (Liver)	9.7	Liver Tissue 6005-N	49.3
Melanoma Metastasis	4.4	Liver Cancer 064003	23.2
Melanoma Margin (Lung)	60.3	Normal Bladder	18.7
Normal Kidney	18.6	Bladder Cancer 1023	15.3
Kidney Ca, Nuclear grade 2 (OD04338)	47.6	Bladder Cancer A302173	15.1
Kidney Margin (OD04338)	26.6	Normal Stomach	58.2
Kidney Ca Nuclear grade 1/2 (OD04339)	49.7	Gastric Cancer 9060397	9.1
Kidney Margin (OD04339)	16.4	Stomach Margin 9060396	33.0
Kidney Ca, Clear cell type (OD04340)	11.9	Gastric Cancer 9060395	29.9
Kidney Margin (OD04340)	19.1	Stomach Margin 9060394	65.5
Kidney Ca, Nuclear grade 3 (OD04348)	7.3	Gastric Cancer 064005	19.5

Table AWF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3015, Run 164043871	Rel. Exp.(%) Ag3070, Run 164525657	Tissue Name	Rel. Exp.(%) Ag3015, Run 164043871	Rel. Exp.(%) Ag3070, Run 164525657
Secondary Th1 act	35.4	35.4	HUVEC IL-1beta	2.6	1.2
Secondary Th2 act	27.9	21.6	HUVEC IFN gamma	9.7	10.5
Secondary Tr1 act	28.9	33.7	HUVEC TNF alpha + IFN gamma	8.1	8.1
Secondary Th1 rest	21.3	21.5	HUVEC TNF alpha + IL4	4.8	7.4
Secondary Th2 rest	24.0	21.2	HUVEC IL-11	4.9	5.6
Secondary Tr1 rest	26.6	26.6	Lung Microvascular EC none	6.9	9.2
Primary Th1 act	23.5	19.6	Lung Microvascular EC TNFalpha + IL-1beta	6.7	7.5
Primary Th2 act	18.3	20.7	Microvascular	12.4	14.8

			Dermal EC none		
Primary Tr1 act	23.8	27.5	Microsvascular Dermal EC TNFalpha + IL- Ibeta	5.5	7.0
Primary Th1 rest	48.6	54.0	Bronchial epithelium TNFalpha + IL1beta	5.1	4.8
Primary Th2 rest	33.0	28.7	Small airway epithelium none	2.0	2.1
Primary Tr1 rest	26.1	28.1	Small airway epithelium TNFalpha + IL- Ibeta	3.4	3.0
CD45RA CD4 lymphocyte act	12.7	14.8	Coronary artery SMC rest	8.4	7.1
CD45RO CD4 lymphocyte act	28.9	32.3	Coronary artery SMC TNFalpha + IL-1beta	4.1	7.3
CD8 lymphocyte act	27.7	42.0	Astrocytes rest	5.3	7.6
Secondary CD8 lymphocyte rest	29.1	35.8	Astrocytes TNFalpha + IL- Ibeta	3.8	3.2
Secondary CD8 lymphocyte act	38.2	41.8	KU-812 (Basophil) rest	4.0	4.5
CD4 lymphocyte none	0.0	18.0	KU-812 (Basophil) PMA/ionomycin	7.1	10.4
2ry Th1/Th2/Tr1_anti- CD95 CH11	26.8	29.7	CCD1106 (Keratinocytes) none	8.3	6.8
LAK cells rest	50.3	71.2	CCD1106 (Keratinocytes) TNFalpha + IL- Ibeta	4.2	2.1
LAK cells IL-2	30.1	35.6	Liver cirrhosis	3.2	3.0
LAK cells IL-2+IL- 12	16.6	29.3	Lupus kidney	3.1	3.3
LAK cells IL- 2+IFN gamma	43.8	44.8	NCI-H292 none	7.5	5.7
LAK cells IL-2+ IL-18	26.8	50.0	NCI-H292 IL-4	6.5	8.3
LAK cells PMA/ionomycin	23.8	24.8	NCI-H292 IL-9	6.0	8.2
NK Cells IL-2 rest	31.4	35.1	NCI-H292 IL-13	3.7	4.1

Two Way MLR 3 day	32.1	50.3	NCI-H292 IFN gamma	3.9	4.5
Two Way MLR 5 day	20.7	26.2	HPAEC none	9.8	10.2
Two Way MLR 7 day	17.1	25.9	HPAEC TNF alpha + IL-1 beta	3.8	4.5
PBMC rest	16.2	19.1	Lung fibroblast none	20.0	23.3
PBMC PWM	55.1	58.6	Lung fibroblast TNF alpha + IL-1 beta	7.6	8.0
PBMC PHA-L	39.8	43.5	Lung fibroblast IL-4	19.9	24.7
Ramos (B cell) none	4.8	5.2	Lung fibroblast IL-9	15.6	19.2
Ramos (B cell) ionomycin	18.9	25.3	Lung fibroblast IL-13	15.1	22.7
B lymphocytes PWM	73.7	77.9	Lung fibroblast IFN gamma	22.8	20.7
B lymphocytes CD40L and IL-4	25.2	37.4	Dermal fibroblast CCD1070 rest	16.4	17.6
EOL-1 dbcAMP	11.7	10.5	Dermal fibroblast CCD1070 TNF alpha	56.6	67.4
EOL-1 dbcAMP PMA/ionomycin	14.7	24.3	Dermal fibroblast CCD1070 IL-1 beta	7.4	9.6
Dendritic cells none	89.5	100.0	Dermal fibroblast IFN gamma	13.1	12.0
Dendritic cells LPS	38.2	55.9	Dermal fibroblast IL-4	24.1	33.7
Dendritic cells anti-CD40	100.0	94.0	IBD Colitis 2	0.9	1.4
Monocytes rest	49.3	72.2	IBD Crohn's	1.4	1.1
Monocytes LPS	15.4	14.7	Colon	10.0	10.7
Macrophages rest	80.7	87.1	Lung	16.5	15.0
Macrophages LPS	24.0	30.1	Thymus	10.3	11.0
HUVEC none	6.9	7.9	Kidney	34.6	50.3
HUVEC starved	12.3	15.3			

CNS_neurodegeneration_v1.0 Summary: Ag3015/Ag2070 This panel does not show differential expression of the NOV54a gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3015/Ag2070 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression in fetal heart (CTs=29-30). This expression is higher than the expression seen in adult heart (CTs=32-33). Thus, expression of this gene could be used to differentiate between the two sources of this tissue. This gene is expressed in other metabolic tissues including adipose, adrenal, liver, pancreas, skeletal muscle and thyroid. This gene encodes a homolog of protein phosphatase 2C (PP2C), which has been linked to the regulation of hormone-sensitive lipase, the rate-limiting enzyme in adipose tissue lipolysis (Eur J Biochem 1987 Oct 15;168(2):399-405). PP2C may also play a role in controlling insulin signaling. Therefore, a therapeutic modulator of this gene and/or gene-product may prove useful in the treatment of diseases affecting the endocrine system.

In addition, protein phosphatase 2C plays a role in dopamine and serotonin signaling. Specifically, PP2C counters the action of these neurotransmitters on DARPP-32. These neurotransmitter systems are the primary targets of drugs that treat schizophrenia and depression. Therefore, agents that inhibit this gene product may have utility in treating these disorders.

References:

Desdouits F, Siciliano JC, Nairn AC, Greengard P, Girault JA. Dephosphorylation of Ser-137 in DARPP-32 by protein phosphatases 2A and 2C: different roles in vitro and in striatonigral neurons. Biochem J 1998 Feb 15;330 (Pt 1):211-6

DARPP-32 (dopamine- and cAMP-regulated phosphoprotein, Mr=32000) is highly expressed in striatonigral neurons in which its phosphorylation is regulated by several neurotransmitters including dopamine and glutamate. DARPP-32 becomes a potent inhibitor of protein phosphatase 1 when it is phosphorylated on Thr-34 by cAMP- or cGMP-dependent protein kinases. DARPP-32 is also phosphorylated on Ser-137 by protein kinase CK1 (CK1), in vitro and in vivo. This phosphorylation has an important regulatory role since it inhibits the dephosphorylation of Thr-34 by calcineurin in vitro and in striatonigral neurons. Here, we show that DARPP-32 phosphorylated by CK1 is a substrate in vitro for protein phosphatases 2A and 2C, but not protein phosphatase 1 or calcineurin. However, in substantia nigra slices, dephosphorylation of Ser-137 was markedly sensitive to decreased temperature, and not detectably affected by the presence of okadaic acid under conditions in which dephosphorylation of Thr-34 by protein phosphatase 2A was inhibited. These results suggest that, in neurons, phospho-Ser-137-DARPP-32 is dephosphorylated by protein phosphatase 2C, but not 2A. Thus, DARPP-32 appears to be a component of a regulatory cascade of

phosphatases in which dephosphorylation of Ser-136 by protein phosphatase 2C facilitates dephosphorylation of Thr-34 by calcineurin, removing the cyclic nucleotide-induced inhibition of protein phosphatase 1.

Overall, expression of this gene is appears to be associated with normal tissues over cancer cell lines. Thus, expression of this gene could be used to differentiate between normal and malignant tissues and potentially as a treatment for cancer.

Panel 2.2 Summary: Ag3070 As seen in the previous panel, the NOV54a gene shows greater expression in normal tissues than in samples derived from malignant tissue. Thus, expression of this gene may be useful in distinguishing the two types of tissue.

Panel 4D Summary: Ag3015/Ag2070 Two experiments with the same probe and primer set produce results that are in excellent agreement. The NOV54a gene, a protein phosphatase 2C homolog is expressed by T lymphocytes prepared under a number of conditions at moderate levels and is expressed at higher levels in treated and untreated dendritic cells, monocytes, and macrophages. Dendritic cells and macrophages are powerful antigen-presenting cells (APC) whose function is pivotal in the initiation and maintenance of normal immune responses. Autoimmunity and inflammation may also be reduced by suppression of this function. Therefore, small molecule drugs that antagonize the function of this gene product may reduce or eliminate the symptoms in patients with several types of autoimmune and inflammatory diseases, such as lupus erythematosus, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, or psoriasis.

NOV55

Expression of gene NOV55 was assessed using the primer-probe set Ag3024, described in Table AXA. Results of the RTQ-PCR runs are shown in Tables AXB, AXC, AXD, AXE, AXF, AXG and AXH.

Table AXA. Probe Name Ag3024

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggtagggctgctataacttgact-3'	22	1033	1207
Probe	TET-5'-tgaaagaaacaccatcctgttgcaga-3'-TAMRA	26	1069	1208
Reverse	5'-tggtcttcaggttggttctttgc-3'	22	1097	1209

Table AXB. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag3024, Run 248122026	Tissue Name	Rel. Exp.(%) Ag3024, Run 248122026
110967 COPD-F	10.6	112427 Match Control Psoriasis-F	37.1
110980 COPD-F	9.5	112418 Psoriasis-M	6.1
110968 COPD-M	10.8	112723 Match Control Psoriasis-M	0.0
110977 COPD-M	16.7	112419 Psoriasis-M	16.6
110989 Emphysema-F	44.8	112424 Match Control Psoriasis-M	17.8
110992 Emphysema-F	3.8	112420 Psoriasis-M	100.0
110993 Emphysema-F	2.6	112425 Match Control Psoriasis-M	26.2
110994 Emphysema-F	1.0	104689 (MF) OA Bone-Backus	0.0
110995 Emphysema-F	9.0	104690 (MF) Adj "Normal" Bone-Backus	0.4
110996 Emphysema-F	1.3	104691 (MF) OA Synovium-Backus	3.0
110997 Asthma-M	1.0	104692 (BA) OA Cartilage-Backus	1.0
111001 Asthma-F	33.4	104694 (BA) OA Bone-Backus	0.0
111002 Asthma-F	37.9	104695 (BA) Adj "Normal" Bone-Backus	0.6
111003 Atopic Asthma-F	41.5	104696 (BA) OA Synovium-Backus	0.5
111004 Atopic Asthma-F	21.2	104700 (SS) OA Bone- Backus	0.2
111005 Atopic Asthma-F	39.0	104701 (SS) Adj "Normal" Bone-Backus	0.3
111006 Atopic Asthma-F	8.5	104702 (SS) OA Synovium-Backus	4.5
111417 Allergy-M	26.6	117093 OA Cartilage Rep7	42.0
112347 Allergy-M	8.6	112672 OA Bone5	18.4
112349 Normal Lung-F	9.2	112673 OA Synovium5	8.0
112357 Normal Lung-F	0.6	112674 OA Synovial Fluid cells5	2.4
112354 Normal Lung-M	1.3	117100 OA Cartilage Rep14	5.5

112374 Crohns-F	4.0	112756 OA Bone9	3.5
112389 Match Control Crohns-F	1.3	112757 OA Synovium9	2.1
112375 Crohns-F	4.5	112758 OA Synovial Fluid Cells9	11.9
112732 Match Control Crohns-F	0.0	117125 RA Cartilage Rep2	2.5
112725 Crohns-M	9.9	113492 Bone2 RA	1.7
112387 Match Control Crohns-M	15.3	113493 Synovium2 RA	0.0
112378 Crohns-M	9.7	113494 Syn Fluid Cells RA	1.6
112390 Match Control Crohns-M	23.7	113499 Cartilage4 RA	0.4
112726 Crohns-M	2.9	113500 Bone4 RA	1.3
112731 Match Control Crohns-M	1.4	113501 Synovium4 RA	0.0
112380 Ulcer Col-F	8.7	113502 Syn Fluid Cells4 RA	0.2
112734 Match Control Ulcer Col-F	0.7	113495 Cartilage3 RA	0.2
112384 Ulcer Col-F	25.3	113496 Bone3 RA	0.6
112737 Match Control Ulcer Col-F	3.0	113497 Synovium3 RA	0.9
112386 Ulcer Col-F	3.3	113498 Syn Fluid Cells3 RA	0.6
112738 Match Control Ulcer Col-F	0.6	117106 Normal Cartilage Rep20	14.1
112381 Ulcer Col-M	1.8	113663 Bone3 Normal	8.5
112735 Match Control Ulcer Col-M	32.8	113664 Synovium3 Normal	4.0
112382 Ulcer Col-M	1.5	113665 Syn Fluid Cells3 Normal	5.7
112394 Match Control Ulcer Col-M	1.8	117107 Normal Cartilage Rep22	13.8
112383 Ulcer Col-M	6.3	113667 Bone4 Normal	17.2
112736 Match Control Ulcer Col-M	1.1	113668 Synovium4 Normal	14.7
112423 Psoriasis-F	12.7	113669 Syn Fluid Cells4 Normal	7.6

Table AX. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3024, Run	Rel. Exp.(%) Ag3024, Run	Tissue Name	Rel. Exp.(%) Ag3024, Run	Rel. Exp.(%) Ag3024, Run
-------------	-----------------------------	-----------------------------	----------------	-----------------------------	-----------------------------

	211011006	233677302		211011006	233677302
AD 1 Hippo	11.7	15.4	Control (Path) 3 Temporal Ctx	1.9	1.9
AD 2 Hippo	23.3	32.3	Control (Path) 4 Temporal Ctx	25.2	41.2
AD 3 Hippo	10.5	13.6	AD 1 Occipital Ctx	7.7	11.7
AD 4 Hippo	6.2	9.9	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	77.9	80.7	AD 3 Occipital Ctx	2.9	4.5
AD 6 Hippo	46.0	45.7	AD 4 Occipital Ctx	16.6	27.2
Control 2 Hippo	28.7	53.6	AD 5 Occipital Ctx	10.9	20.3
Control 4 Hippo	2.8	2.4	AD 6 Occipital Ctx	61.1	51.4
Control (Path) 3 Hippo	4.5	3.8	Control 1 Occipital Ctx	1.3	1.1
AD 1 Temporal Ctx	4.6	7.0	Control 2 Occipital Ctx	52.1	72.7
AD 2 Temporal Ctx	16.8	26.1	Control 3 Occipital Ctx	9.9	13.1
AD 3 Temporal Ctx	2.9	3.8	Control 4 Occipital Ctx	1.6	1.4
AD 4 Temporal Ctx	15.8	19.1	Control (Path) 1 Occipital Ctx	81.2	87.1
AD 5 Inf Temporal Ctx	47.6	94.6	Control (Path) 2 Occipital Ctx	12.2	16.5

AD 5 SupTemporal Ctx	42.9	46.0	Control (Path) 3 Occipital Ctx	0.6	0.6
AD 6 Inf Temporal Ctx	22.8	29.9	Control (Path) 4 Occipital Ctx	11.0	14.3
AD 6 Sup Temporal Ctx	24.0	30.1	Control 1 Parietal Ctx	3.5	4.0
Control 1 Temporal Ctx	2.3	2.8	Control 2 Parietal Ctx	16.3	16.8
Control 2 Temporal Ctx	35.8	48.0	Control 3 Parietal Ctx	17.2	26.2
Control 3 Temporal Ctx	13.1	14.5	Control (Path) 1 Parietal Ctx	100.0	100.0
Control 4 Temporal Ctx	4.4	4.2	Control (Path) 2 Parietal Ctx	29.3	38.7
Control (Path) 1 Temporal Ctx	56.3	68.8	Control (Path) 3 Parietal Ctx	1.1	2.4
Control (Path) 2 Temporal Ctx	47.6	54.0	Control (Path) 4 Parietal Ctx	42.0	56.6

Table AXD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3024, Run 165517896	Tissue Name	Rel. Exp.(%) Ag3024, Run 165517896
Liver adenocarcinoma	0.4	Kidney (fetal)	0.2
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	2.4
Adrenal gland	2.6	Renal ca. RXF 393	0.0
Thyroid	0.7	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	2.0	Renal ca. TK-10	0.0
Brain (fetal)	11.0	Liver	0.7
Brain (whole)	54.7	Liver (fetal)	0.0
Brain (amygdala)	27.2	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	12.9	Lung	0.0
Brain (hippocampus)	36.3	Lung (fetal)	0.0
Brain (substantia nigra)	5.7	Lung ca. (small cell)	0.0

		LX-1	
Brain (thalamus)	64.2	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	100.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	3.8	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.2
glio/astro U-118-MG	14.7	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	1.7	Lung ca. (non-s.cell) HOP-62	0.6
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	2.1
astrocytoma SNB-75	6.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.6	Mammary gland	0.4
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	11.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.9
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	2.3
Heart	0.6	Breast ca. BT-549	1.0
Skeletal muscle (fetal)	1.9	Breast ca. MDA-N	0.0
Skeletal muscle	0.4	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.0
Thymus	0.0	Ovarian ca. OVCAR- 4	0.0
Spleen	0.7	Ovarian ca. OVCAR- 5	0.2
Lymph node	0.0	Ovarian ca. OVCAR- 8	0.0
Colorectal	0.8	Ovarian ca. IGROV- 1	0.0
Stomach	6.9	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	1.4	Uterus	2.2
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.4

Tissue Name	Rel. Exp.(%) Ag3024, Run 163577593	Tissue Name	Rel. Exp.(%) Ag3024, Run 163577593
Normal Colon	4.7	Kidney Margin 8120608	1.3
CC Well to Mod Diff (ODO3866)	5.0	Kidney Cancer 8120613	100.0
CC Margin (ODO3866)	3.1	Kidney Margin 8120614	5.3
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	0.8
CC Margin (ODO3868)	0.3	Kidney Margin 9010321	6.3
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.7
CC Margin (ODO3920)	0.5	Uterus Cancer 064011	9.4
CC Gr.2 ascend colon (ODO3921)	3.1	Normal Thyroid	5.7
CC Margin (ODO3921)	2.1	Thyroid Cancer 064010	0.3
CC from Partial Hepatectomy (ODO4309) Mets	1.8	Thyroid Cancer A302152	2.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	23.5
Colon mets to lung (OD04451-01)	0.0	Normal Breast	7.9
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	7.4

Normal Prostate 6546-1	3.7	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	5.2	Breast Cancer Mets (OD04590-03)	3.6
Prostate Margin (OD04410)	5.4	Breast Cancer Metastasis (OD04655-05)	3.8
Prostate Cancer (OD04720-01)	11.9	Breast Cancer 064006	7.1
Prostate Margin (OD04720-02)	6.4	Breast Cancer 1024	6.2
Normal Lung 061010	0.0	Breast Cancer 9100266	27.7
Lung Met to Muscle (ODO4286)	0.6	Breast Margin 9100265	4.3
Muscle Margin (ODO4286)	1.0	Breast Cancer A209073	30.8
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	3.0
Lung Margin (OD03126)	1.6	Normal Liver	1.3
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	0.4
Lung Margin (OD04404)	1.0	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	8.4	Liver Tissue 6004-N	1.1
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	2.5	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	2.8
Melanoma Mets to Lung (OD04321)	0.4	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.5	Bladder Cancer A302173	2.2
Normal Kidney	23.7	Bladder Cancer (OD04718-01)	0.7
Kidney Ca, Nuclear grade 2 (OD04338)	1.2	Bladder Normal Adjacent (OD04718-03)	0.4
Kidney Margin (OD04338)	10.6	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	1.4
Kidney Margin (OD04339)	7.7	Ovarian Cancer (OD04768-07)	0.0

Kidney Ca, Clear cell type (OD04340)	9.3	Ovary Margin (OD04768-08)	3.3
Kidney Margin (OD04340)	17.2	Normal Stomach	17.8
Kidney Ca, Nuclear grade 3 (OD04348)	2.3	Gastric Cancer 9060358	1.0
Kidney Margin (OD04348)	5.1	Stomach Margin 9060359	13.6
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	7.5
Kidney Margin (OD04622-03)	2.8	Stomach Margin 9060394	4.1
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	13.4
Kidney Margin (OD04450-03)	17.8	Stomach Margin 9060396	6.7
Kidney Cancer 8120607	1.7	Gastric Cancer 064005	1.2

Table AXF, Panel 3D

Tissue Name	Rel. Exp.(%) Ag3024, Run 164886426	Tissue Name	Rel. Exp.(%) Ag3024, Run 164886426
Daoy- Medulloblastoma	1.6	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.3
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.6	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.6
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.4	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	14.1	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.2	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.4	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	15.3	JM1- pre-B-cell lymphoma	0.0
Cerebellum	5.4	Jurkat- T cell leukemia	0.2
NCI-H292-	0.0	TF-1- Erythroleukemia	0.0

Mucoepidermoid lung carcinoma			
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	3.3	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	57.4	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	14.4	Caki-2- Clear cell renal carcinoma	32.8
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.2
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	2.9	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.3	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	15.9	SU86.86- Pancreatic carcinoma (liver metastasis)	1.5
NCI-UMC-11- Lung carcinoid	1.2	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.3
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.8
NCI-H716- Colon cancer	100.0	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	0.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	0.0	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	1.1	HT-1080- Fibrosarcoma	0.3

NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	4.6
NCI-SNU-16- Gastric carcinoma	0.4	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.2
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	1.3
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.2	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table AXG. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3024, Run 162427416	Tissue Name	Rel. Exp.(%) Ag3024, Run 162427416
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	31.4
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	14.9
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	8.8
Secondary Th2 rest	4.4	HUVEC IL-11	11.4
Secondary Tr1 rest	0.0	Lung Microvascular EC none	10.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	8.7
Primary Th2 act	0.0	Microvascular Dermal EC none	11.3
Primary Tr1 act	0.0	Microsvascular Dermal EC TNFalpha + IL-1beta	3.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	5.3
Primary Th2 rest	0.0	Small airway epithelium none	3.7

Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	66.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	21.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	10.4
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	17.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	5.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	3.2
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	10.4
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	6.3
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	3.2
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	4.7	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	4.1
Dendritic cells none	0.0	Dermal fibroblast IFN	6.1

		gamma	
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	8.7
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	7.7
Monocytes LPS	0.0	Colon	24.1
Macrophages rest	0.0	Lung	14.5
Macrophages LPS	5.3	Thymus	100.0
HUVEC none	24.7	Kidney	0.0
HUVEC starved	11.8		

Table AXH. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag3024, Run 171694538	Tissue Name	Rel. Exp.(%) Ag3024, Run 171694538
BA4 Control	38.2	BA17 PSP	27.7
BA4 Control2	73.2	BA17 PSP2	10.7
BA4 Alzheimer's2	7.7	Sub Nigra Control	11.7
BA4 Parkinson's	60.7	Sub Nigra Control2	20.4
BA4 Parkinson's2	85.3	Sub Nigra Alzheimer's2	6.4
BA4 Huntington's	44.1	Sub Nigra Parkinson's2	29.5
BA4 Huntington's2	8.7	Sub Nigra Huntington's	28.9
BA4 PSP	7.7	Sub Nigra Huntington's2	20.7
BA4 PSP2	39.5	Sub Nigra PSP2	1.5
BA4 Depression	2.1	Sub Nigra Depression	0.8
BA4 Depression2	5.8	Sub Nigra Depression2	4.7
BA7 Control	66.0	Glob Palladus Control	3.9
BA7 Control2	52.5	Glob Palladus Control2	6.9
BA7 Alzheimer's2	8.8	Glob Palladus Alzheimer's	5.5
BA7 Parkinson's	20.4	Glob Palladus Alzheimer's2	2.8
BA7 Parkinson's2	64.6	Glob Palladus Parkinson's	48.6
BA7 Huntington's	55.1	Glob Palladus Parkinson's2	7.1

BA7 Huntington's2	54.7	Glob Palladus PSP	3.2
BA7 PSP	57.8	Glob Palladus PSP2	3.0
BA7 PSP2	26.4	Glob Palladus Depression	2.1
BA7 Depression	7.7	Temp Pole Control	14.8
BA9 Control	31.4	Temp Pole Control2	57.0
BA9 Control2	100.0	Temp Pole Alzheimer's	6.1
BA9 Alzheimer's	4.8	Temp Pole Alzheimer's2	3.4
BA9 Alzheimer's2	14.5	Temp Pole Parkinson's	23.2
BA9 Parkinson's	31.6	Temp Pole Parkinson's2	30.1
BA9 Parkinson's2	56.3	Temp Pole Huntington's	41.8
BA9 Huntington's	53.2	Temp Pole PSP	3.5
BA9 Huntington's2	14.3	Temp Pole PSP2	3.3
BA9 PSP	17.7	Temp Pole Depression2	1.9
BA9 PSP2	5.1	Cing Gyr Control	67.8
BA9 Depression	9.0	Cing Gyr Control2	38.4
BA9 Depression2	7.4	Cing Gyr Alzheimer's	7.4
BA17 Control	51.8	Cing Gyr Alzheimer's2	12.2
BA17 Control2	54.0	Cing Gyr Parkinson's	15.4
BA17 Alzheimer's2	2.7	Cing Gyr Parkinson's2	39.5
BA17 Parkinson's	23.2	Cing Gyr Huntington's	48.0
BA17 Parkinson's2	53.6	Cing Gyr Huntington's2	14.2
BA17 Huntington's	41.2	Cing Gyr PSP	13.9
BA17 Huntington's2	9.5	Cing Gyr PSP2	5.0
BA17 Depression	6.0	Cing Gyr Depression	5.7
BA17 Depression2	14.4	Cing Gyr Depression2	9.7

AI_comprehensive panel_v1.0 Summary: Ag3024 The NOV55 gene is found at low but significant levels in lung tissue from COPD, emphysema and asthma patients. This expression is consistent with panel 4D which shows expression in small airway epithelium. Therefore, this gene could be a marker or a target for lung inflammatory diseases.

5 **CNS_neurodegeneration_v1.0 Summary:** Ag3024 Results of two experiments with the same probe and primer set confirm expression of the NOV55 gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3024 Expression of the NOV55 gene, a heparin sulfate proteoglycan homolog, is highly brain preferential and suggests a role for this gene product in
10 CNS processes. Heparin sulfate proteoglycans (HSPGs) are a component of amyloid plaques in Alzheimer's disease. The interaction of apoE with HSPGs has also been implicated in the pathogenesis of Alzheimer's disease and may play a role in neuronal repair. apoE has an HSPG-binding site highly complementary to heparan sulfates rich in N- and O-sulfo groups. Therefore, enzymes that influence the structure of HSPGs, such as the putative protein product
15 of the NOV55 gene, may influence protein aggregation and the functional processes underlying Alzheimer's disease. Thus, agents that target and modulate the activity of this gene product may be effective in the treatment of neurodegenerative diseases including Alzheimer's disease. This gene is also expressed in breast and brain cancer cell lines at low but significant levels. Therefore, the expression of this gene could be of use as a marker for breast and brain cancer.
20 In addition, therapeutic inhibition of the activity of the product of this gene, through the use of antibodies or small molecule drugs, may be useful in the therapy of brain and breast cancer.

References:

Libeu CP, Lund-Katz S, Phillips MC, Wehrli S, Hernaiz MJ, Capila I, Linhardt RJ, Raffai RL, Newhouse YM, Zhou F, Weisgraber KH. New insights into the heparan sulfate
25 proteoglycan-binding activity of apolipoprotein E. J Biol Chem 2001 Oct 19;276(42):39138-44

Defective binding of apolipoprotein E (apoE) to heparan sulfate proteoglycans (HSPGs) is associated with increased risk of atherosclerosis due to inefficient clearance of lipoprotein remnants by the liver. The interaction of apoE with HSPGs has also been
30 implicated in the pathogenesis of Alzheimer's disease and may play a role in neuronal repair. To identify which residues in the heparin-binding site of apoE and which structural elements of heparan sulfate interact, we used a variety of approaches, including glycosaminoglycan specificity assays, (13)C nuclear magnetic resonance, and heparin affinity chromatography. The formation of the high affinity complex required Arg-142, Lys-143, Arg-145, Lys-146, and

Arg-147 from apoE and N- and 6-O-sulfo groups of the glucosamine units from the heparin fragment. As shown by molecular modeling, using a high affinity binding octasaccharide fragment of heparin, these findings are consistent with a binding mode in which five
 5 saccharide residues of fully sulfated heparan sulfate lie in a shallow groove of the alpha-helix that contains the HSPG-binding site (helix 4 of the four-helix bundle of the 22-kDa fragment). This groove is lined with residues Arg-136, Ser-139, His-140, Arg-142, Lys-143, Arg-145, Lys-146, and Arg-147. In the model, all of these residues make direct contact with either the 2-O-sulfo groups of the iduronic acid monosaccharides or the N- and 6-O-sulfo groups of the glucosamine sulfate monosaccharides. This model indicates that apoE has an HSPG-binding
 10 site highly complementary to heparan sulfate rich in N- and O-sulfo groups such as that found in the liver and the brain.

Inoue S. Basement membrane and beta amyloid fibrillogenesis in Alzheimer's disease. Int Rev Cytol 2001;210:121-61

High-resolution ultrastructural and immunohistochemical studies revealed that in situ
 15 beta amyloid fibrils of Alzheimer's disease were made up of a core consisting of a solid column of amyloid P component (AP) and associated chondroitin sulfate proteoglycan, and a heparan sulfate proteoglycan surface layer with externally associated fine filaments of beta protein. The main body of beta amyloid fibrils closely resembled that of microfibrils. Abundant microfibrils were reported to be present at the basement membrane of capillaries
 20 with "leaky" blood-urine or blood-air barriers. Similarly, abundant microfibril-like beta amyloid fibrils are formed at the microvascular basement membrane in cerebrovascular amyloid angiopathy with altered blood-brain barrier. Since AP is an indispensable major component of microfibrils and microfibril-like structures, the formation of microfibrils may depend on, among other factors, the availability of AP. Thus, in beta amyloid fibrillogenesis
 25 fibrils may be built around AP which continuously leaks out from circulation into vascular basement membrane, and beta amyloid fibrils may be regarded as pathologically altered basement membrane-associated microfibrils. With no source of AP around them, senile plaque fibrils may also be derived from perivascular amyloid.

Panel 2D Summary: Ag3024 The NOV55 gene is expressed at low but significant
 30 levels in most of the samples on this panel, with highest expression in a kidney cancer sample (CT=30.6). Significant levels of expression are also seen in samples derived from breast and gastric cancer samples.

Therefore, expression of this gene could be of use as a marker for breast and gastric cancer. In addition, therapeutic inhibition of the activity of the product of this gene, through

the use of antibodies or small molecule drugs, may be useful in the therapy of breast and gastric cancer.

Panel 3D Summary: Ag3024 The NOV55 gene is expressed at low but significant levels in cell lines from a renal carcinoma, colon cancer, glioblastoma and three lung cancer lines. Thus, this gene could be a marker as well as a target for inhibition in these cancers.

Panel 4D Summary: Ag3024 The NOV55 gene, a heparin Sulfate 6-Sulfotransferase 3 homolog, is expressed at low but significant levels in thymus and small airway epithelium treated with TNFalpha + IL-1beta (CTs=34). Thus, the NOV55 gene product may be a marker for thymus or activated small airway epithelium.

Panel CNS_1 Summary: Ag3024 Expression of the NOV55 gene in this panel confirms the presence of this gene product in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

NOV56a and NOV56b

Expression of gene NOV56a and variant NOV56b was assessed using the primer-probe sets Ag3027 and Ag1169, described in Tables AYA and AYB. Results of the RTQ-PCR runs are shown in Tables AYC, AYD, AYE, AYF, AYG, AYH and AYI.

Table AYA. Probe Name Ag3027

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaaaccagatttggagttcgtt-3'	22	355	1210
Probe	TET-5'-cttgaaatgtcctcaccacaactgat-3'-TAMRA	26	377	1211
Reverse	5'-tccagatagatggtggaatcag-3'	22	425	1212

Table AYB. Probe Name Ag1169

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaaaccagatttggagttcgtt-3'	22	355	1213
Probe	TET-5'-cttgaaatgtcctcaccacaactgat-3'-TAMRA	26	377	1214
Reverse	5'-tccagatagatggtggaatcag-3'	22	425	1215

Table AYC. General_screening_panel_v1.5

Tissue Name	Rel. Exp.(%) Ag3027, Run 228714682	Tissue Name	Rel. Exp.(%) Ag3027, Run 228714682
Adipose	2.3	Renal ca. TK-10	10.4
Melanoma*	1.4	Bladder	11.0

Hs688(A).T			
Melanoma* Hs688(B).T	1.0	Gastric ca. (liver met.) NCI-N87	0.6
Melanoma* M14	0.0	Gastric ca. KATO III	7.4
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	43.8
Melanoma* SK- MEL-5	1.9	Colon ca. SW480	6.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	9.9
Testis Pool	0.7	Colon ca. HT29	2.0
Prostate ca.* (bone met) PC-3	0.7	Colon ca. HCT-116	0.5
Prostate Pool	3.8	Colon ca. CaCo-2	73.2
Placenta	0.0	Colon cancer tissue	21.8
Uterus Pool	0.4	Colon ca. SW1116	0.7
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	2.6
Ovarian ca. SK-OV- 3	29.9	Colon ca. SW-48	0.3
Ovarian ca. OVCAR-4	0.8	Colon Pool	2.6
Ovarian ca. OVCAR-5	11.0	Small Intestine Pool	0.8
Ovarian ca. IGROV- 1	6.9	Stomach Pool	2.0
Ovarian ca. OVCAR-8	1.1	Bone Marrow Pool	0.0
Ovary	0.8	Fetal Heart	1.4
Breast ca. MCF-7	0.2	Heart Pool	0.5
Breast ca. MDA- MB-231	0.5	Lymph Node Pool	1.4
Breast ca. BT 549	0.5	Fetal Skeletal Muscle	1.1
Breast ca. T47D	0.4	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	4.9	Thymus Pool	0.3
Trachea	0.3	CNS cancer (glio/astro) U87-MG	4.7
Lung	0.0	CNS cancer (glio/astro) U-118-MG	1.6
Fetal Lung	2.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	0.5
Lung ca. LX-1	100.0	CNS cancer (astro)	4.1

		SNB-75	
Lung ca. NCI-H146	0.3	CNS cancer (glio) SNB-19	13.8
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	3.6
Lung ca. A549	1.4	Brain (Amygdala) Pool	0.2
Lung ca. NCI-H526	1.0	Brain (cerebellum)	0.1
Lung ca. NCI-H23	2.6	Brain (fetal)	0.2
Lung ca. NCI-H460	1.2	Brain (Hippocampus) Pool	0.3
Lung ca. HOP-62	0.2	Cerebral Cortex Pool	0.5
Lung ca. NCI-H522	0.7	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.4
Fetal Liver	3.2	Brain (whole)	0.6
Liver ca. HepG2	0.5	Spinal Cord Pool	0.3
Kidney Pool	1.8	Adrenal Gland	1.1
Fetal Kidney	29.9	Pituitary gland Pool	0.3
Renal ca. 786-0	1.8	Salivary Gland	0.2
Renal ca. A498	0.7	Thyroid (female)	0.1
Renal ca. ACHN	3.2	Pancreatic ca. CAPAN2	3.1
Renal ca. UO-31	3.2	Pancreas Pool	4.4

Table AYD. Panel 1.2

Tissue Name	Rel. Exp.(%) Ag1169, Run 129128191	Rel. Exp.(%) Ag1169, Run 129656838	Tissue Name	Rel. Exp.(%) Ag1169, Run 129128191	Rel. Exp.(%) Ag1169, Run 129656838
Endothelial cells	0.0	24.3	Renal ca. 786-0	0.2	10.5
Heart (Fetal)	0.0	57.0	Renal ca. A498	5.5	10.9
Pancreas	0.0	0.0	Renal ca. RXF393	0.9	14.9
Pancreatic ca. CAPAN 2	0.2	0.0	Renal ca. ACHN	1.5	22.8
Adrenal Gland	0.0	7.0	Renal ca. UO-31	5.6	12.3
Thyroid	0.1	0.1	Renal ca. TK-10	11.3	1.7
Salivary gland	0.1	6.2	Liver	0.0	12.9
Pituitary gland	0.0	0.3	Liver (fetal)	0.0	1.4
Brain (fetal)	0.0	0.0	Liver ca. (hepatoblast) HepG2	0.2	4.5

Brain (whole)	0.0	4.5	Lung	0.0	14.9
Brain (amygdala)	0.0	6.3	Lung (fetal)	0.0	22.2
Brain (cerebellum)	0.0	0.3	Lung ca. (small cell) LX-1	27.0	74.2
Brain (hippocampus)	0.0	14.4	Lung ca. (small cell) NCI-H69	0.0	0.0
Brain (thalamus)	0.0	1.2	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Cerebral Cortex	0.0	15.5	Lung ca. (large cell) NCI-H460	0.0	0.0
Spinal cord	0.0	22.4	Lung ca. (non-sm. cell) A549	0.9	0.0
glio/astro U87-MG	0.2	0.0	Lung ca. (non-s.cell) NCI-H23	0.3	0.0
glio/astro U-118-MG	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.1	0.6
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	0.1	0.1
neuro*; met SK-N-AS	0.0	0.0	Lung ca. (squam.) SW 900	6.3	0.0
astrocytoma SF-539	0.0	0.0	Lung ca. (squam.) NCI-H596	0.0	0.0
astrocytoma SNB-75	0.0	0.0	Mammary gland	100.0	100.0
glioma SNB-19	0.4	0.6	Breast ca.* (pl.ef) MCF-7	0.0	4.0
glioma U251	0.1	1.7	Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
glioma SF-295	0.5	0.0	Breast ca.* (pl.ef) T47D	0.0	1.4
Heart	0.0	58.2	Breast ca. BT-549	0.0	0.1
Skeletal Muscle	0.0	9.3	Breast ca. MDA-N	0.0	0.0
Bone marrow	0.0	0.0	Ovary	0.0	14.0
Thymus	0.0	0.0	Ovarian ca. OVCAR-3	0.0	5.2
Spleen	0.0	0.0	Ovarian ca. OVCAR-4	0.0	15.3

Lymph node	0.0	3.0	Ovarian ca. OVCAR-5	50.7	15.7
Colorectal Tissue	0.1	0.0	Ovarian ca. OVCAR-8	1.5	0.1
Stomach	8.4	23.0	Ovarian ca. IGROV-1	43.5	17.0
Small intestine	0.2	10.0	Ovarian ca. (ascites) SK-OV-3	13.5	17.0
Colon ca. SW480	0.6	0.0	Uterus	0.2	7.7
Colon ca.* SW620 (SW480 met)	7.5	0.4	Placenta	0.0	75.8
Colon ca. HT29	0.0	0.0	Prostate	12.1	13.7
Colon ca. HCT-116	0.0	0.0	Prostate ca.* (bone met) PC-3	0.0	10.4
Colon ca. CaCo-2	34.4	11.8	Testis	0.6	0.2
Colon ca. Tissue (ODO3866)	11.4	15.9	Melanoma Hs688(A).T	0.0	0.0
Colon ca. HCC-2998	2.0	0.0	Melanoma* (met) Hs688(B).T	0.2	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma UACC-62	0.0	0.0
Bladder	5.3	11.3	Melanoma M14	0.0	0.0
Trachea	0.0	2.6	Melanoma LOX IMVI	0.0	0.0
Kidney	4.5	14.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney (fetal)	14.3	31.2			

Table AYE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1169, Run 165518394	Rel. Exp.(%) Ag3027, Run 165519993	Tissue Name	Rel. Exp.(%) Ag1169, Run 165518394	Rel. Exp.(%) Ag3027, Run 165519993
Liver adenocarcinoma	0.5	0.4	Kidney (fetal)	0.7	0.0
Pancreas	0.0	0.0	Renal ca. 786-0	0.5	0.6

Pancreatic ca. CAPAN 2	0.7	0.4	Renal ca. A498	0.9	1.6
Adrenal gland	0.0	0.0	Renal ca. RXF 393	0.0	0.9
Thyroid	0.0	0.0	Renal ca. ACHN	0.0	0.0
Salivary gland	0.0	0.0	Renal ca. UO- 31	0.0	0.0
Pituitary gland	0.0	0.0	Renal ca. TK- 10	2.1	0.4
Brain (fetal)	0.0	0.0	Liver	0.0	0.0
Brain (whole)	0.0	0.0	Liver (fetal)	0.0	0.0
Brain (amygdala)	0.0	0.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	0.0	0.0	Lung	0.0	0.0
Brain (hippocampus)	0.0	0.4	Lung (fetal)	0.0	0.0
Brain (substantia nigra)	0.3	0.0	Lung ca. (small cell) LX-1	21.5	12.0
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.0
Cerebral Cortex	0.0	0.5	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	0.0	0.3	Lung ca. (large cell)NCI-H460	0.0	1.3
glio/astro U87-MG	0.0	0.3	Lung ca. (non- sm. cell) A549	0.0	0.0
glio/astro U-118- MG	0.0	0.3	Lung ca. (non- s.cell) NCI- H23	0.7	0.3
astrocytoma SW1783	0.0	0.0	Lung ca. (non- s.cell) HOP-62	0.0	0.0
neuro*; met SK-N- AS	0.0	0.0	Lung ca. (non- s.cl) NCI- H522	0.0	0.0
astrocytoma SF- 539	0.3	0.0	Lung ca. (squam.) SW 900	2.8	0.0
astrocytoma SNB- 75	1.9	0.9	Lung ca. (squam.) NCI- H596	0.0	0.0
glioma SNB-19	0.0	0.4	Mammary gland	100.0	100.0

glioma U251	0.4	0.4	Breast ca.* (pl.ef) MCF-7	0.5	0.0
glioma SF-295	0.0	0.0	Breast ca.* (pl.ef) MDA- MB-231	0.0	0.3
Heart (fetal)	0.0	0.0	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	0.0	0.0	Breast ca. BT- 549	0.3	0.0
Skeletal muscle (fetal)	0.0	0.0	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.0	0.0	Ovary	0.0	0.0
Bone marrow	0.0	0.0	Ovarian ca. OVCA-3	0.0	0.0
Thymus	0.9	0.0	Ovarian ca. OVCA-4	0.4	0.0
Spleen	0.7	0.0	Ovarian ca. OVCA-5	0.9	1.4
Lymph node	0.0	0.0	Ovarian ca. OVCA-8	0.0	0.0
Colorectal	0.9	0.2	Ovarian ca. IGROV-1	1.0	2.3
Stomach	1.0	0.6	Ovarian ca.* (ascites) SK- OV-3	1.4	1.8
Small intestine	0.7	0.0	Uterus	0.5	0.0
Colon ca. SW480	0.8	0.4	Placenta	0.0	0.0
Colon ca.* SW620(SW480 met)	0.6	0.4	Prostate	0.3	0.5
Colon ca. HT29	0.0	0.8	Prostate ca.* (bone met)PC- 3	0.0	0.0
Colon ca. HCT- 116	0.9	0.3	Testis	0.8	1.2
Colon ca. CaCo-2	6.7	8.1	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	2.9	2.0	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC- 2998	0.4	0.3	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.3	0.0	Melanoma M14	0.0	0.0
Bladder	1.7	0.4	Melanoma LOX IMVI	0.0	0.0

Trachea	0.0	0.0	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	2.7	1.3	Adipose	0.7	0.0

Table AYF. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3027, Run 163577594	Tissue Name	Rel. Exp.(%) Ag3027, Run 163577594
Normal Colon	14.6	Kidney Margin 8120608	5.9
CC Well to Mod Diff (ODO3866)	15.1	Kidney Cancer 8120613	34.4
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	5.7
CC Gr.2 rectosigmoid (ODO3868)	15.5	Kidney Cancer 9010320	10.5
CC Margin (ODO3868)	0.7	Kidney Margin 9010321	25.0
CC Mod Diff (ODO3920)	64.2	Normal Uterus	0.0
CC Margin (ODO3920)	2.2	Uterus Cancer 064011	18.4
CC Gr.2 ascend colon (ODO3921)	11.2	Normal Thyroid	1.6
CC Margin (ODO3921)	1.8	Thyroid Cancer 064010	1.6
CC from Partial Hepatectomy (ODO4309) Mets	16.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.3	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	6.8	Normal Breast	28.7
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	0.5
Normal Prostate 6546-1	14.9	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	29.7	Breast Cancer Mets (OD04590-03)	3.7
Prostate Margin (OD04410)	19.1	Breast Cancer Metastasis (OD04655-05)	1.7
Prostate Cancer (OD04720-01)	4.5	Breast Cancer 064006	0.4
Prostate Margin (OD04720-02)	3.0	Breast Cancer 1024	55.1

Normal Lung 061010	0.5	Breast Cancer 9100266	1.9
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	8.2
Muscle Margin (ODO4286)	0.9	Breast Cancer A209073	18.3
Lung Malignant Cancer (OD03126)	11.0	Breast Margin A2090734	34.6
Lung Margin (OD03126)	1.1	Normal Liver	0.0
Lung Cancer (OD04404)	1.8	Liver Cancer 064003	15.0
Lung Margin (OD04404)	0.4	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.5	Liver Cancer 1026	0.0
Lung Margin (OD04565)	2.1	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	3.3	Liver Tissue 6004-N	1.1
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	2.6
Ocular Mel Met to Liver (ODO4310)	1.6	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	13.5
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	32.5
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	2.6
Normal Kidney	14.6	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	3.3	Bladder Normal Adjacent (OD04718-03)	1.0
Kidney Margin (OD04338)	5.1	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	7.9	Ovarian Cancer 064008	1.8
Kidney Margin (OD04339)	27.0	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	3.4	Ovary Margin (OD04768-08)	0.5
Kidney Margin (OD04340)	61.6	Normal Stomach	45.7
Kidney Ca, Nuclear grade 3 (OD04348)	0.4	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	84.1	Stomach Margin 9060359	8.2
Kidney Cancer (OD04622-01)	3.5	Gastric Cancer 9060395	6.4

Kidney Margin (OD04622-03)	1.4	Stomach Margin 9060394	12.9
Kidney Cancer (OD04450-01)	39.2	Gastric Cancer 9060397	42.9
Kidney Margin (OD04450-03)	51.8	Stomach Margin 9060396	31.4
Kidney Cancer 8120607	4.0	Gastric Cancer 064005	100.0

Table AYG. Panel 3D

Tissue Name	Rel. Exp.(%) Ag1169, Run 164038616	Tissue Name	Rel. Exp.(%) Ag1169, Run 164038616
Daoy- Medulloblastoma	1.9	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	0.0	JM1- pre-B-cell lymphoma	0.0
Cerebellum	0.0	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	1.8	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	100.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	2.6	769-P- Clear cell renal carcinoma	1.0

NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	2.3
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	3.7
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	3.5
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	2.3
NCI-UMC-11- Lung carcinoid	1.7	BxPC-3- Pancreatic adenocarcinoma	1.2
LX-1- Small cell lung cancer	20.6	HPAC- Pancreatic adenocarcinoma	3.3
Colo-205- Colon cancer	0.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	12.3	CFPAC-1- Pancreatic ductal adenocarcinoma	2.6
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	6.0	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	2.6	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	25.5	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	10.7	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	1.0
NCI-SNU-5- Gastric carcinoma	1.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	0.0	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	2.0
NCI-SNU-1- Gastric carcinoma	29.7	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0

RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	3.5	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table AYH. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1169, Run 139591349	Rel. Exp.(%) Ag1169, Run 145735616	Rel. Exp.(%) Ag3027, Run 162426723	Tissue Name	Rel. Exp.(%) Ag1169, Run 139591349	Rel. Exp.(%) Ag1169, Run 145735616	Rel. Exp.(%) Ag3027, Run 162426723
Secondary Th1 act	0.0	0.0	0.0	HUVEC IL-1beta	0.0	0.0	0.0
Secondary Th2 act	0.0	0.0	0.0	HUVEC IFN gamma	0.0	0.0	0.7
Secondary Tr1 act	0.0	0.0	0.0	HUVEC TNF alpha + IFN gamma	0.0	0.0	0.0
Secondary Th1 rest	0.0	0.0	0.0	HUVEC TNF alpha + IL4	0.4	0.0	0.0
Secondary Th2 rest	0.0	0.0	0.0	HUVEC IL-11	0.0	0.0	0.0
Secondary Tr1 rest	0.0	0.0	0.0	Lung Microvascular EC none	0.4	0.0	0.0
Primary Th1 act	0.0	0.0	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.9	0.0	0.0
Primary Th2 act	0.0	0.0	0.0	Microvascular Dermal EC none	0.5	0.0	3.1

Primary Tr1 act	0.0	0.0	0.0	Microvas ular Dermal EC TNFalpha + IL-1beta	0.0	0.0	0.0
Primary Th1 rest	0.0	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	1.0	0.0	5.1
Primary Th2 rest	0.0	0.0	0.0	Small airway epithelium none	0.0	0.0	2.0
Primary Tr1 rest	0.0	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0	1.3	1.9
CD45RA CD4 lymphocyt e act	0.0	0.0	0.0	Coronary artery SMC rest	0.0	0.0	0.0
CD45RO CD4 lymphocyt e act	0.0	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0	0.0
CD8 lymphocyt e act	0.0	0.0	0.0	Astrocytes rest	0.0	0.0	0.0
Secondary CD8 lymphocyt e rest	0.0	0.0	0.0	Astrocytes TNFalpha + IL-1beta	1.4	3.1	0.6
Secondary CD8 lymphocyt e act	0.0	0.0	0.0	KU-812 (Basophil) rest	0.0	0.0	0.0
CD4 lymphocyt e none	0.0	0.0	0.0	KU-812 (Basophil) PMA/iono mycin	0.0	0.0	0.0
2ry Th1/Th2/T r1_anti- CD95 CH11	0.0	0.0	0.0	CCD1106 (Keratinoc ytes) none	0.0	0.0	0.0
LAK cells rest	0.0	0.0	0.0	CCD1106 (Keratinoc	0.0	0.0	0.0

				ytes) TNFalpha + IL-1beta			
LAK cells IL-2	0.0	0.0	0.0	Liver cirrhosis	2.1	1.3	3.2
LAK cells IL-2+IL- 12	0.0	0.0	0.0	Lupus kidney	5.3	3.7	4.8
LAK cells IL-2+IFN gamma	0.0	0.0	0.0	NCI-H292 none	0.9	0.0	2.2
LAK cells IL-2+ IL- 18	0.0	0.0	0.0	NCI-H292 IL-4	0.0	1.2	0.7
LAK cells PMA/iono mycin	0.0	0.0	0.0	NCI-H292 IL-9	0.3	0.0	0.9
NK Cells IL-2 rest	0.0	0.0	0.0	NCI-H292 IL-13	0.0	0.0	0.7
Two Way MLR 3 day	0.0	0.0	0.0	NCI-H292 IFN gamma	0.0	3.7	0.0
Two Way MLR 5 day	0.0	0.0	0.0	HPAEC none	36.1	0.0	1.1
Two Way MLR 7 day	0.0	0.0	1.1	HPAEC TNF alpha + IL-1 beta	0.0	0.0	0.0
PBMC rest	0.0	0.0	0.0	Lung fibroblast none	0.0	0.0	0.0
PBMC PWM	0.0	1.4	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0	0.0	0.0
PBMC PHA-L	0.5	0.0	1.0	Lung fibroblast IL-4	0.0	0.0	0.0
Ramos (B cell) none	0.0	1.8	0.0	Lung fibroblast IL-9	0.0	0.0	2.0
Ramos (B cell) ionomycin	0.5	0.0	0.0	Lung fibroblast IL-13	0.0	0.0	1.0
B lymphocyt es PWM	0.0	0.0	1.3	Lung fibroblast IFN gamma	0.5	0.0	0.0

B lymphocytes CD40L and IL-4	0.0	0.0	0.0	Dermal fibroblast CCD1070 rest	0.0	0.0	0.0
EOL-1 dbcAMP	0.0	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0	0.0	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	0.0	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.3	0.0	0.0
Dendritic cells none	2.1	3.5	1.1	Dermal fibroblast IFN gamma	0.0	0.0	0.0
Dendritic cells LPS	0.0	0.0	0.0	Dermal fibroblast IL-4	0.0	0.0	0.0
Dendritic cells anti-CD40	0.0	0.0	0.0	IBD Colitis 2	1.3	3.6	2.4
Monocytes rest	0.0	0.0	0.0	IBD Crohn's	0.4	0.0	1.0
Monocytes LPS	0.0	0.0	0.0	Colon	8.1	32.3	56.6
Macrophages rest	0.0	0.0	4.8	Lung	0.5	0.0	0.0
Macrophages LPS	0.0	0.0	1.8	Thymus	100.0	100.0	100.0
HUVEC none	0.0	0.0	0.0	Kidney	0.9	0.0	0.0
HUVEC starved	0.0	0.0	1.3				

Table AYI. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag3027, Run 225051163	Tissue Name	Rel. Exp.(%) Ag3027, Run 225051163
97457_Patient-02go_adipose	0.0	94709_Donor 2 AM - A_adipose	9.5
97476_Patient-07sk_skeletal muscle	3.2	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	15.0	94711_Donor 2 AM - C_adipose	5.3

1204

seen in samples derived from colon cancer and ovarian cancer. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of lung, colon, and breast cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung, colon and breast cancers.

While expression of this gene is seen predominantly in cancer cell lines, significant expression is also seen in fetal kidney (CT=31.8). Furthermore, expression is higher in fetal kidney than in adult kidney (CT=35.8). Thus, expression of this gene could be used to differentiate between adult and fetal kidney. In addition, the expression in fetal kidney suggests that this gene product may be involved in the development of the kidney. Therefore, therapeutic modulation of the expression or function of this gene may be useful in treating disease of the kidney.

Panel 1.2 Summary: Ag1169 Results from one experiment, Run 129128191, with the NOV56a gene are in agreement with Results in Panel 1.3D and General_screening_panel_v1.5. A second run, Run 129656838, produces disparate results.

Panel 1.3D Summary: Ag1169/Ag3027 Two experiments with the same probe and primer both show highest expression of the gene NOV56a in the mammary gland (CTs=31). Low, but significant levels of expression are also seen in a lung cancer cell line. Thus, expression of this gene may be used to differentiate between these samples and other samples on this panel.

Panel 2D Summary: Ag3027 Highest expression of the NOV56a gene is seen in a gastric cancer. Significant expression is also seen in breast cancer, colon cancer and normal kidney. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker for the presence of breast, colon and kidney cancers. A second experiment with the probe/primer set Ag1169 is not included. The amp plot indicates that there were experimental difficulties with this run.

Panel 3D Summary: Ag1169 Expression of the NOV56a gene is restricted to samples derived from lung and gastric cancer cell lines (CTs=33-35). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of lung and gastric cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung and gastric cancers.

Panel 4D Summary: Ag1169/Ag3027 Two experiments with the same probe and primer set show expression of the NOV56a gene limited to the thymus (CTs=32-33). Thus,

expression of this gene could be used as a marker for thymic tissue. Furthermore, this restricted expression suggests that this gene product may play an important role in T cell development. Therefore, small molecule therapeutics, or antibody therapeutics designed against the protein encoded for by this gene could be utilized to modulate immune function (T cell development) and be important for organ transplant, AIDS treatment or post chemotherapy immune reconstitution.

Panel 5 Islet Summary: Ag3027 Expression of the NOV56a gene is restricted to a sample derived from the kidney (CT=34.9). This expression is consistent with expression in Panel 1.3D. Thus, expression of this gene could be used as a marker for kidney tissue.

10 NOV57

Expression of gene NOV57 was assessed using the primer-probe sets Ag3031, Ag1301b and Ag1415, described in Tables AZA, AZB and AZC. Results of the RTQ-PCR runs are shown in Tables AZD, AZE, AZF and AZG.

Table AZA. Probe Name Ag3031

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaaagggtgatgtctggagcat-3'	21	616	1216
Probe	TET-5'-tgtatgtcatgctctgtgccagccta-3'-TAMRA	26	648	1217
Reverse	5'-gatgtctgtgtcgtcaaaagg-3'	21	674	1218

15

Table AZB. Probe Name Ag1301b

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaaagggtgatgtctggagcat-3'	21	616	1219
Probe	TET-5'-tgtatgtcatgctctgtgccagccta-3'-TAMRA	26	648	1220
Reverse	5'-gatgtctgtgtcgtcaaaagg-3'	21	674	1221

Table AZC. Probe Name Ag1415

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aaaagggtgatgtctggagcat-3'	21	616	1222
Probe	TET-5'-tgtatgtcatgctctgtgccagccta-3'-TAMRA	26	648	1223
Reverse	5'-gatgtctgtgtcgtcaaaagg-3'	21	674	1224

Table AZD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3031, Run 211011868	Rel. Exp.(%) Ag3031, Run 225437445	Tissue Name	Rel. Exp.(%) Ag3031, Run 211011868	Rel. Exp.(%) Ag3031, Run 225437445
AD 1 Hippo	23.8	46.7	Control (Path) 3 Temporal Ctx	6.0	10.2
AD 2 Hippo	21.2	21.9	Control (Path) 4 Temporal Ctx	26.4	28.1
AD 3 Hippo	14.2	11.6	AD 1 Occipital Ctx	31.9	41.8
AD 4 Hippo	7.4	10.5	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	94.0	94.0	AD 3 Occipital Ctx	17.9	17.4
AD 6 Hippo	53.6	61.6	AD 4 Occipital Ctx	12.9	17.3
Control 2 Hippo	12.7	17.0	AD 5 Occipital Ctx	35.1	20.9
Control 4 Hippo	14.6	15.2	AD 6 Occipital Ctx	20.9	27.5
Control (Path) 3 Hippo	22.2	7.2	Control 1 Occipital Ctx	4.9	4.9
AD 1 Temporal Ctx	42.3	42.9	Control 2 Occipital Ctx	46.7	32.1
AD 2 Temporal Ctx	78.5	29.9	Control 3 Occipital Ctx	17.8	19.2
AD 3 Temporal Ctx	17.9	20.4	Control 4 Occipital Ctx	13.2	12.1
AD 4 Temporal Ctx	30.1	31.2	Control (Path) 1 Occipital Ctx	49.3	59.9

AD 5 Inf Temporal Ctx	100.0	94.6	Control (Path) 2 Occipital Ctx	13.6	8.8
AD 5 SupTemporal Ctx	59.9	52.5	Control (Path) 3 Occipital Ctx	6.1	4.7
AD 6 Inf Temporal Ctx	75.3	95.9	Control (Path) 4 Occipital Ctx	31.2	21.2
AD 6 Sup Temporal Ctx	94.6	100.0	Control 1 Parietal Ctx	10.3	11.1
Control 1 Temporal Ctx	8.5	7.3	Control 2 Parietal Ctx	56.6	60.7
Control 2 Temporal Ctx	16.3	18.8	Control 3 Parietal Ctx	15.4	15.0
Control 3 Temporal Ctx	13.0	15.3	Control (Path) 1 Parietal Ctx	34.2	41.8
Control 4 Temporal Ctx	12.8	16.6	Control (Path) 2 Parietal Ctx	15.4	20.3
Control (Path) 1 Temporal Ctx	26.8	31.0	Control (Path) 3 Parietal Ctx	8.4	7.0
Control (Path) 2 Temporal Ctx	25.5	17.8	Control (Path) 4 Parietal Ctx	44.1	31.6

Table AZE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag1301b, Run 165528224	Rel. Exp.(%) Ag3031, Run 167961982	Tissue Name	Rel. Exp.(%) Ag1301b, Run 165528224	Rel. Exp.(%) Ag3031, Run 167961982
Liver adenocarcinoma	18.4	32.3	Kidney (fetal)	10.7	66.0
Pancreas	4.7	11.9	Renal ca. 786-0	15.3	25.5
Pancreatic ca. CAPAN 2	4.9	4.9	Renal ca. A498	11.9	12.5
Adrenal gland	23.0	11.0	Renal ca. RXF 393	14.9	18.0
Thyroid	7.1	8.5	Renal ca. ACHN	8.0	10.5

Salivary gland	10.1	13.7	Renal ca. UO-31	6.0	1.1
Pituitary gland	18.2	9.2	Renal ca. TK-10	7.4	18.6
Brain (fetal)	24.0	42.3	Liver	11.9	17.8
Brain (whole)	43.8	19.6	Liver (fetal)	20.6	14.7
Brain (amygdala)	20.4	9.6	Liver ca. (hepatoblast) HepG2	22.2	28.7
Brain (cerebellum)	50.7	50.3	Lung	17.0	22.8
Brain (hippocampus)	18.7	8.1	Lung (fetal)	10.4	26.4
Brain (substantia nigra)	8.4	8.9	Lung ca. (small cell) LX-1	9.1	17.4
Brain (thalamus)	20.0	10.6	Lung ca. (small cell) NCI-H69	1.5	4.0
Cerebral Cortex	5.3	7.2	Lung ca. (s.cell var.) SHP-77	18.4	57.4
Spinal cord	9.6	7.0	Lung ca. (large cell) NCI-H460	50.3	4.2
glio/astro U87-MG	12.1	16.5	Lung ca. (non-sm. cell) A549	8.0	28.1
glio/astro U-118-MG	15.5	13.4	Lung ca. (non-s.cell) NCI-H23	10.2	18.2
astrocytoma SW1783	6.9	11.1	Lung ca. (non-s.cell) HOP-62	20.3	33.0
neuro*; met SK-N-AS	11.4	10.9	Lung ca. (non-s.cl) NCI-H522	9.7	23.5
astrocytoma SF-539	12.5	25.5	Lung ca. (squam.) SW 900	4.0	8.5
astrocytoma SNB-75	10.9	18.9	Lung ca. (squam.) NCI-H596	1.6	4.8
glioma SNB-19	32.1	25.3	Mammary gland	36.9	22.5
glioma U251	100.0	100.0	Breast ca.* (pl.ef) MCF-7	5.4	6.7
glioma SF-295	18.4	46.0	Breast ca.* (pl.ef) MDA-MB-231	16.7	7.9

Heart (fetal)	2.3	6.2	Breast ca.* (pl.ef) T47D	13.2	46.7
Heart	8.4	8.2	Breast ca. BT- 549	9.9	6.6
Skeletal muscle (fetal)	6.3	24.1	Breast ca. MDA-N	1.6	11.9
Skeletal muscle	20.2	7.9	Ovary	3.0	5.1
Bone marrow	21.8	25.2	Ovarian ca. OVCAR-3	5.1	12.0
Thymus	18.2	38.4	Ovarian ca. OVCAR-4	4.7	8.8
Spleen	26.6	14.0	Ovarian ca. OVCAR-5	8.3	42.3
Lymph node	42.6	23.7	Ovarian ca. OVCAR-8	2.8	0.9
Colorectal	16.8	16.8	Ovarian ca. IGROV-1	2.6	11.7
Stomach	37.4	5.9	Ovarian ca.* (ascites) SK- OV-3	10.0	29.9
Small intestine	36.6	14.2	Uterus	38.4	17.4
Colon ca. SW480	3.7	4.5	Placenta	11.1	9.4
Colon ca.* SW620(SW480 met)	3.8	24.0	Prostate	18.6	8.5
Colon ca. HT29	1.6	4.6	Prostate ca.* (bone met)PC- 3	6.9	14.1
Colon ca. HCT- 116	3.0	9.9	Testis	45.1	26.6
Colon ca. CaCo-2	2.1	5.1	Melanoma Hs688(A).T	0.8	1.2
Colon ca. tissue(ODO3866)	5.4	3.6	Melanoma* (met) Hs688(B).T	3.8	3.4
Colon ca. HCC- 2998	6.2	15.6	Melanoma UACC-62	4.6	8.0
Gastric ca.* (liver met) NCI-N87	20.0	13.3	Melanoma M14	28.5	9.9
Bladder	8.4	12.2	Melanoma LOX IMVI	0.7	8.1
Trachea	13.6	6.4	Melanoma* (met) SK- MEL-5	2.2	2.9
Kidney	10.4	23.8	Adipose	8.1	12.3

Table AZF. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag1301b, Run 173859869	Tissue Name	Rel. Exp.(%) Ag1301b, Run 173859869
Normal Colon	39.2	Kidney Margin (OD04348)	79.0
Colon cancer (OD06064)	6.3	Kidney malignant cancer (OD06204B)	6.0
Colon Margin (OD06064)	13.0	Kidney normal adjacent tissue (OD06204E)	15.4
Colon cancer (OD06159)	0.0	Kidney Cancer (OD04450-01)	24.5
Colon Margin (OD06159)	27.0	Kidney Margin (OD04450-03)	23.7
Colon cancer (OD06297-04)	2.5	Kidney Cancer 8120613	3.0
Colon Margin (OD06297-015)	39.8	Kidney Margin 8120614	38.2
CC Gr.2 ascend colon (ODO3921)	3.8	Kidney Cancer 9010320	8.0
CC Margin (ODO3921)	4.4	Kidney Margin 9010321	10.8
Colon cancer metastasis (OD06104)	14.3	Kidney Cancer 8120607	15.2
Lung Margin (OD06104)	17.2	Kidney Margin 8120608	3.8
Colon mets to lung (OD04451-01)	0.0	Normal Uterus	33.7
Lung Margin (OD04451-02)	10.2	Uterine Cancer 064011	33.9
Normal Prostate	30.6	Normal Thyroid	2.9
Prostate Cancer (OD04410)	7.6	Thyroid Cancer 064010	4.4
Prostate Margin (OD04410)	25.5	Thyroid Cancer A302152	29.9
Normal Ovary	25.0	Thyroid Margin A302153	7.3
Ovarian cancer (OD06283-03)	8.1	Normal Breast	37.6
Ovarian Margin (OD06283-07)	34.4	Breast Cancer (OD04566)	21.6
Ovarian Cancer 064008	46.3	Breast Cancer 1024	100.0
Ovarian cancer (OD06145)	34.4	Breast Cancer (OD04590-01)	25.2
Ovarian Margin (OD06145)	52.9	Breast Cancer Mets (OD04590-03)	35.6

Ovarian cancer (OD06455-03)	11.6	Breast Cancer Metastasis (OD04655-05)	45.1
Ovarian Margin (OD06455-07)	18.8	Breast Cancer 064006	21.2
Normal Lung	22.8	Breast Cancer 9100266	26.6
Invasive poor diff. lung adeno (ODO4945-01)	18.7	Breast Margin 9100265	19.3
Lung Margin (ODO4945-03)	13.9	Breast Cancer A209073	6.1
Lung Malignant Cancer (OD03126)	11.7	Breast Margin A2090734	35.8
Lung Margin (OD03126)	12.2	Breast cancer (OD06083)	49.3
Lung Cancer (OD05014A)	9.0	Breast cancer node metastasis (OD06083)	25.9
Lung Margin (OD05014B)	35.4	Normal Liver	36.3
Lung cancer (OD06081)	23.7	Liver Cancer 1026	2.5
Lung Margin (OD06081)	27.9	Liver Cancer 1025	45.7
Lung Cancer (OD04237-01)	13.3	Liver Cancer 6004-T	30.1
Lung Margin (OD04237-02)	28.5	Liver Tissue 6004-N	27.7
Ocular Melanoma Metastasis	11.5	Liver Cancer 6005-T	6.7
Ocular Melanoma Margin (Liver)	27.0	Liver Tissue 6005-N	17.4
Melanoma Metastasis	21.6	Liver Cancer 064003	32.3
Melanoma Margin (Lung)	14.7	Normal Bladder	13.2
Normal Kidney	17.8	Bladder Cancer 1023	23.0
Kidney Ca, Nuclear grade 2 (OD04338)	50.3	Bladder Cancer A302173	20.0
Kidney Margin (OD04338)	24.3	Normal Stomach	89.5
Kidney Ca Nuclear grade 1/2 (OD04339)	62.0	Gastric Cancer 9060397	5.7
Kidney Margin (OD04339)	16.3	Stomach Margin 9060396	17.7
Kidney Ca, Clear cell type (OD04340)	7.3	Gastric Cancer 9060395	19.6
Kidney Margin (OD04340)	10.7	Stomach Margin 9060394	42.6
Kidney Ca, Nuclear	4.6	Gastric Cancer 064005	7.5

grade 3 (OD04348)			
-------------------	--	--	--

Table AZG. Panel 4D

Tissue Name	Rel. Exp.(%) Ag1301b, Run 138983163	Rel. Exp.(%) Ag1415, Run 138642033	Rel. Exp.(%) Ag3031, Run 162426783	Tissue Name	Rel. Exp.(%) Ag1301b, Run 138983163	Rel. Exp.(%) Ag1415, Run 138642033	Rel. Exp.(%) Ag3031, Run 162426783
Secondary Th1 act	17.0	25.3	13.8	HUVEC IL-1beta	11.5	13.0	11.3
Secondary Th2 act	22.7	18.9	20.0	HUVEC IFN gamma	28.3	24.7	20.2
Secondary Tr1 act	32.3	26.4	20.0	HUVEC TNF alpha + IFN gamma	3.2	6.6	12.3
Secondary Th1 rest	22.2	20.3	12.0	HUVEC TNF alpha + IL4	11.7	9.7	9.7
Secondary Th2 rest	42.6	37.9	21.0	HUVEC IL-11	10.1	8.2	11.6
Secondary Tr1 rest	27.9	27.5	26.8	Lung Microvascular EC none	33.9	28.7	25.2
Primary Th1 act	33.4	29.9	19.1	Lung Microvascular EC TNFalpha + IL-1beta	23.2	19.1	24.8
Primary Th2 act	28.1	41.2	13.3	Microvascular Dermal EC none	41.8	45.1	25.7
Primary Tr1 act	49.0	52.5	20.9	Microvascular Dermal EC TNFalpha + IL-1beta	27.5	40.3	22.4
Primary Th1 rest	80.7	87.1	80.7	Bronchial epithelium TNFalpha + IL1beta	20.3	32.3	22.1
Primary Th2 rest	67.4	68.8	64.2	Small airway epithelium	7.3	3.3	4.8

				none			
Primary Tr1 rest	46.0	50.3	54.7	Small airway epithelium TNFalpha + IL-1beta	34.4	35.1	26.8
CD45RA CD4 lymphocyt e act	12.9	8.0	16.5	Coronary artery SMC rest	7.9	7.9	12.3
CD45RO CD4 lymphocyt e act	33.4	44.8	22.4	Coronary artery SMC TNFalpha + IL-1beta	7.6	10.7	8.1
CD8 lymphocyt e act	22.8	23.0	18.4	Astrocytes rest	7.9	10.3	12.7
Secondary CD8 lymphocyt e rest	22.2	25.5	25.0	Astrocytes TNFalpha + IL-1beta	8.7	5.8	13.3
Secondary CD8 lymphocyt e act	19.3	21.8	22.5	KU-812 (Basophil) rest	40.1	35.4	48.0
CD4 lymphocyt e none	41.5	42.3	34.6	KU-812 (Basophil) PMA/iono mycin	57.8	61.1	78.5
2ry Th1/Th2/T r1_anti- CD95 CH11	65.1	54.7	41.2	CCD1106 (Keratinoc ytes) none	3.8	8.0	6.6
LAK cells rest	28.7	37.1	32.3	CCD1106 (Keratinoc ytes) TNFalpha + IL-1beta	27.4	24.5	5.3
LAK cells IL-2	38.7	49.0	36.3	Liver cirrhosis	17.0	9.4	5.0
LAK cells IL-2+IL- 12	26.8	27.4	26.8	Lupus kidney	24.1	23.8	9.5
LAK cells IL-2+IFN gamma	43.5	45.4	42.9	NCI-H292 none	38.7	49.3	45.4

LAK cells IL-2+ IL-18	26.2	25.2	35.6	NCI-H292 IL-4	58.6	51.4	46.7
LAK cells PMA/iono mycin	8.3	8.6	3.3	NCI-H292 IL-9	56.3	46.0	54.3
NK Cells IL-2 rest	28.7	35.8	32.8	NCI-H292 IL-13	30.4	31.9	23.3
Two Way MLR 3 day	42.3	49.0	46.3	NCI-H292 IFN gamma	16.7	28.3	29.1
Two Way MLR 5 day	17.9	16.2	13.5	HPAEC none	15.7	26.4	19.1
Two Way MLR 7 day	14.7	12.6	14.4	HPAEC TNF alpha + IL-1 beta	23.8	32.5	27.9
PBMC rest	21.9	29.9	19.2	Lung fibroblast none	13.0	11.0	11.9
PBMC PWM	66.9	53.2	50.3	Lung fibroblast TNF alpha + IL-1 beta	8.7	7.2	13.2
PBMC PHA-L	35.6	46.7	23.3	Lung fibroblast IL-4	5.6	12.0	10.5
Ramos (B cell) none	23.5	33.2	16.5	Lung fibroblast IL-9	8.5	7.4	15.3
Ramos (B cell) ionomycin	53.2	53.2	49.3	Lung fibroblast IL-13	20.3	16.3	8.9
B lymphocyt es PWM	29.9	36.3	34.2	Lung fibroblast IFN gamma	11.1	10.4	13.0
B lymphocyt es CD40L and IL-4	33.2	36.6	35.8	Dermal fibroblast CCD1070 rest	47.0	11.3	13.8
EOL-1 dbcAMP	18.9	14.8	12.8	Dermal fibroblast CCD1070 TNF alpha	45.7	53.6	55.1
EOL-1 dbcAMP PMA/iono	30.4	29.1	25.5	Dermal fibroblast CCD1070	16.8	17.3	15.4

mycin				IL-1 beta			
Dendritic cells none	14.9	10.0	13.7	Dermal fibroblast IFN gamma	5.8	8.8	6.5
Dendritic cells LPS	15.7	5.8	8.7	Dermal fibroblast IL-4	17.6	20.0	13.6
Dendritic cells anti-CD40	12.7	17.7	15.2	IBD Colitis 2	2.0	1.3	1.1
Monocytes rest	35.6	27.5	36.6	IBD Crohn's	3.0	1.4	1.4
Monocytes LPS	34.6	43.8	25.0	Colon	33.0	26.2	34.6
Macrophages rest	19.2	16.6	17.2	Lung	6.8	13.5	7.5
Macrophages LPS	17.9	16.2	6.7	Thymus	100.0	100.0	55.5
HUVEC none	9.8	13.2	17.3	Kidney	87.1	79.0	100.0
HUVEC starved	27.2	27.5	28.3				

CNS_neurodegeneration_v1.0 Summary: Ag3031 Two experiments with the same probe and primer set produce results that are in excellent agreement. The NOV57 gene, a kinase homolog, is expressed more highly in the temporal cortex of brains from Alzheimer's disease patients than in the temporal cortex of normal brains unaffected by Alzheimer's disease. Kinases have been shown to play a role in the pathogenesis of Alzheimer's disease. The dysregulation of this kinase, NOV57, indicates an active role for this pathway in disease pathogenesis. Thus, inhibitors of this gene product, by modulating this pathway, may have utility in the treatment of Alzheimer's disease and other neurodegenerative diseases.

References:

Morishima Y, Gotoh Y, Zieg J, Barrett T, Takano H, Flavell R, Davis RJ, Shirasaki Y, Greenberg ME. Beta-amyloid induces neuronal apoptosis via a mechanism that involves the c-Jun N-terminal kinase pathway and the induction of Fas ligand. *J Neurosci* 2001 Oct 1;21(19):7551-60

Elevated levels of beta-Amyloid (Abeta) are present in the brains of individuals with either the sporadic or familial form of Alzheimer's disease (AD), and the deposition of Abeta within the senile plaques that are a hallmark of AD is thought to be a primary cause of the

cognitive dysfunction that occurs in AD. Recent evidence suggests that Abeta induces neuronal apoptosis in the brain and in primary neuronal cultures, and that this Abeta-induced neuronal death may be responsible in part for the cognitive decline found in AD patients. In this study we have characterized one mechanism by which Abeta induces neuronal death. We found that in cortical neurons exposed to Abeta, activated c-Jun N-terminal kinase (JNK) is required for the phosphorylation and activation of the c-Jun transcription factor, which in turn stimulates the transcription of several key target genes, including the death inducer Fas ligand. The binding of Fas ligand to its receptor Fas then induces a cascade of events that lead to caspase activation and ultimately cell death. By analyzing the effects of mutations in each of the components of the JNK-c-Jun-Fas ligand-Fas pathway, we demonstrate that this pathway plays a critical role in mediating Abeta-induced death of cultured neurons. These findings raise the possibility that the JNK pathway may also contribute to Abeta-dependent death in AD patients

Panel 1.3D Summary: Ag1301b/Ag3031 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of the NOV57 gene in a brain cancer cell line (CTs=29-30). Overall, this gene is expressed at moderate to low levels in all the samples in this panel.

This gene has low to moderate expression in several endocrine/metabolic-related tissues, including adipose, pancreas, liver, skeletal muscle and thyroid. Thus, a therapeutic modulator to this gene and/or gene-product may be useful in the treatment of diseases which affect the endocrine system.

Panel 2.2 Summary: Ag1301b The NOV57 gene is expressed in breast cancer at a moderate level. It is also expressed at a higher level in normal gastric, prostate and colon tissues compared to the adjacent tumors. Hence, inhibition of this drug might be used for treatment of breast cancer. It could also be used as a diagnostic marker for gastric, prostate and colon cancers.

Panel 4D Summary: Ag1301b/Ag1415/Ag3031 Three experiments with the same probe and primer sets produce results that are in excellent agreement, with highest expression of the NOV57 gene in the thymus and kidney. This gene is also expressed at higher levels in resting Th1 and Th2 lymphocytes than in activated Th1 and Th2 lymphocytes. Therefore, small molecule agonists of the gene product may be useful as therapeutics to reduce the activation of Th1 and Th2 cells and thus reduce symptoms in patients with autoimmune and inflammatory diseases, such as Crohn's disease, ulcerative colitis, multiple sclerosis, chronic

obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

NOV58a and NOV58b: Gap Junction Beta-5 (connexin)

5 Expression of gene NOV58a and variant NOV58b was assessed using the primer-probe set Ag2914, described in Table BAA.

Table BAA. Probe Name Ag2914

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aacactgtggactgcttcattc-3'	22	517	1225
Probe	TET-5'-ccaaacccactgagaagacgatcttca-3'-TAMRA	27	539	1226
Reverse	5'-atacacaagcatgaggtgatga-3'	22	578	1227

CNS_neurodegeneration_v1.0 Summary: Ag2914 The amp plot indicates that there are experimental difficulties with this run (data not shown).

10 **Panel 1.3D Summary:** Ag2914 Expression of this gene is low/undetectable (CTs >35) across all of the samples on this panel (data not shown).

Panel 2D Summary: Ag2914 Expression of this gene is low/undetectable (CTs >35) across all of the samples on this panel (data not shown).

15 **Panel 4D Summary:** Ag2914 The amp plot indicates that there are experimental difficulties with this run (data not shown).

BB. CG56633-01: TRANSLATION INITIATION FACTOR 5

Expression of gene CG56633-01 was assessed using the primer-probe set Ag2900, described in Table BBA. Results of the RTQ-PCR runs are shown in Tables BBB, BBC, BBD and BBE.

20 **Table BBA. Probe Name Ag2900**

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gctaagttccttgatgcttctg-3'	22	184	1228
Probe	TET-5'-caaaacttgattaccgtcgatgtgca-3'-TAMRA	26	209	1229
Reverse	5'-ccaccagaatgtcaaagagtgt-3'	22	238	1230

Table BBB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2900,	Tissue Name	Rel. Exp.(%) Ag2900,
-------------	----------------------	-------------	----------------------

	Run 206485415		Run 206485415
AD 1 Hippo	12.9	Control (Path) 3 Temporal Ctx	7.0
AD 2 Hippo	33.9	Control (Path) 4 Temporal Ctx	18.6
AD 3 Hippo	7.4	AD 1 Occipital Ctx	14.5
AD 4 Hippo	11.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	68.8	AD 3 Occipital Ctx	8.0
AD 6 Hippo	61.6	AD 4 Occipital Ctx	16.6
Control 2 Hippo	36.1	AD 5 Occipital Ctx	36.3
Control 4 Hippo	21.0	AD 6 Occipital Ctx	25.7
Control (Path) 3 Hippo	7.7	Control 1 Occipital Ctx	6.8
AD 1 Temporal Ctx	19.6	Control 2 Occipital Ctx	65.5
AD 2 Temporal Ctx	27.5	Control 3 Occipital Ctx	14.6
AD 3 Temporal Ctx	6.2	Control 4 Occipital Ctx	12.3
AD 4 Temporal Ctx	17.6	Control (Path) 1 Occipital Ctx	100.0
AD 5 Inf Temporal Ctx	59.0	Control (Path) 2 Occipital Ctx	6.5
AD 5 Sup Temporal Ctx	42.0	Control (Path) 3 Occipital Ctx	5.1
AD 6 Inf Temporal Ctx	65.1	Control (Path) 4 Occipital Ctx	13.2
AD 6 Sup Temporal Ctx	48.3	Control 1 Parietal Ctx	8.0
Control 1 Temporal Ctx	8.0	Control 2 Parietal Ctx	24.5
Control 2 Temporal Ctx	52.9	Control 3 Parietal Ctx	18.3
Control 3 Temporal Ctx	18.2	Control (Path) 1 Parietal Ctx	86.5
Control 3 Temporal Ctx	9.0	Control (Path) 2 Parietal Ctx	17.4
Control (Path) 1 Temporal Ctx	57.0	Control (Path) 3 Parietal Ctx	9.4
Control (Path) 2 Temporal Ctx	38.2	Control (Path) 4 Parietal Ctx	31.6

Tissue Name	Rel. Exp.(%) Ag2900, Run 159996755	Tissue Name	Rel. Exp.(%) Ag2900, Run 159996755
Liver adenocarcinoma	22.7	Kidney (fetal)	14.1
Pancreas	1.8	Renal ca. 786-0	15.1
Pancreatic ca. CAPAN 2	2.8	Renal ca. A498	23.7
Adrenal gland	22.1	Renal ca. RXF 393	2.7
Thyroid	3.6	Renal ca. ACHN	3.0
Salivary gland	2.9	Renal ca. UO-31	6.7
Pituitary gland	5.8	Renal ca. TK-10	3.6
Brain (fetal)	10.7	Liver	6.0
Brain (whole)	12.3	Liver (fetal)	12.5
Brain (amygdala)	8.4	Liver ca. (hepatoblast) HepG2	11.8
Brain (cerebellum)	9.3	Lung	15.0
Brain (hippocampus)	61.6	Lung (fetal)	6.9
Brain (substantia nigra)	5.1	Lung ca. (small cell) LX-1	11.1
Brain (thalamus)	10.4	Lung ca. (small cell) NCI-H69	12.3
Cerebral Cortex	22.4	Lung ca. (s.cell var.) SHP-77	30.4
Spinal cord	4.7	Lung ca. (large cell) NCI-H460	28.3
glio/astro U87-MG	26.1	Lung ca. (non-sm. cell) A549	15.7
glio/astro U-118-MG	94.0	Lung ca. (non-s.cell) NCI-H23	15.0
astrocytoma SW1783	16.5	Lung ca. (non-s.cell) HOP-62	4.2
neuro*; met SK-N-AS	50.0	Lung ca. (non-s.cl) NCI-H522	9.3
astrocytoma SF-539	9.5	Lung ca. (squam.) SW 900	8.4
astrocytoma SNB-75	10.9	Lung ca. (squam.) NCI-H596	2.8
glioma SNB-19	10.2	Mammary gland	10.1
glioma U251	3.3	Breast ca.* (pl.ef) MCF-7	21.8
glioma SF-295	2.9	Breast ca.* (pl.ef) MDA-MB-231	100.0
Heart (fetal)	2.1	Breast ca.* (pl.ef) T47D	5.3
Heart	3.1	Breast ca. BT-549	40.9

Skeletal muscle (fetal)	3.0	Breast ca. MDA-N	12.0
Skeletal muscle	2.1	Ovary	3.0
Bone marrow	8.1	Ovarian ca. OVCAR-3	7.3
Thymus	4.0	Ovarian ca. OVCAR-4	0.9
Spleen	5.3	Ovarian ca. OVCAR-5	5.9
Lymph node	4.0	Ovarian ca. OVCAR-8	8.3
Colorectal	7.6	Ovarian ca. IGROV-1	2.1
Stomach	2.7	Ovarian ca.* (ascites) SK-OV-3	11.3
Small intestine	5.7	Uterus	3.0
Colon ca. SW480	11.0	Placenta	12.6
Colon ca.* SW620(SW480 met)	7.2	Prostate	4.3
Colon ca. HT29	7.7	Prostate ca.* (bone met)PC-3	24.7
Colon ca. HCT-116	23.0	Testis	7.7
Colon ca. CaCo-2	15.7	Melanoma Hs688(A).T	11.3
Colon ca. tissue(ODO3866)	12.8	Melanoma* (met) Hs688(B).T	5.3
Colon ca. HCC-2998	33.4	Melanoma UACC-62	4.2
Gastric ca.* (liver met) NCI-N87	21.9	Melanoma M14	4.4
Bladder	11.8	Melanoma LOX IMVI	28.7
Trachea	10.2	Melanoma* (met) SK-MEL-5	17.9
Kidney	4.3	Adipose	13.1

Table BBD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2900, Run 159996787	Tissue Name	Rel. Exp.(%) Ag2900, Run 159996787
Normal Colon	71.2	Kidney Margin 8120608	3.0
CC Well to Mod Diff (ODO3866)	24.5	Kidney Cancer 8120613	2.0
CC Margin (ODO3866)	20.6	Kidney Margin 8120614	2.5

CC Gr.2 rectosigmoid (ODO3868)	47.3	Kidney Cancer 9010320	7.1
CC Margin (ODO3868)	6.0	Kidney Margin 9010321	5.0
CC Mod Diff (ODO3920)	39.2	Normal Uterus	5.3
CC Margin (ODO3920)	21.0	Uterus Cancer 064011	16.5
CC Gr.2 ascend colon (ODO3921)	69.3	Normal Thyroid	8.2
CC Margin (ODO3921)	16.0	Thyroid Cancer 064010	15.0
CC from Partial Hepatectomy (ODO4309) Mets	41.8	Thyroid Cancer A302152	7.7
Liver Margin (ODO4309)	40.6	Thyroid Margin A302153	11.0
Colon mets to lung (OD04451-01)	14.3	Normal Breast	12.7
Lung Margin (OD04451-02)	15.0	Breast Cancer (OD04566)	21.3
Normal Prostate 6546-1	11.7	Breast Cancer (OD04590-01)	63.7
Prostate Cancer (OD04410)	40.1	Breast Cancer Mets (OD04590-03)	48.0
Prostate Margin (OD04410)	32.3	Breast Cancer Metastasis (OD04655-05)	32.1
Prostate Cancer (OD04720-01)	26.4	Breast Cancer 064006	21.8
Prostate Margin (OD04720-02)	35.8	Breast Cancer 1024	6.3
Normal Lung 061010	46.7	Breast Cancer 9100266	39.0
Lung Met to Muscle (ODO4286)	34.6	Breast Margin 9100265	14.4
Muscle Margin (ODO4286)	12.4	Breast Cancer A209073	37.1
Lung Malignant Cancer (OD03126)	18.8	Breast Margin A2090734	14.4
Lung Margin (OD03126)	16.7	Normal Liver	14.1
Lung Cancer (OD04404)	20.9	Liver Cancer 064003	20.2
Lung Margin (OD04404)	15.3	Liver Cancer 1025	7.8
Lung Cancer (OD04565)	15.8	Liver Cancer 1026	5.6
Lung Margin (OD04565)	10.3	Liver Cancer 6004-T	6.0
Lung Cancer (OD04237-01)	29.1	Liver Tissue 6004-N	6.6

Lung Margin (OD04237-02)	29.5	Liver Cancer 6005-T	6.6
Ocular Mel Met to Liver (ODO4310)	17.0	Liver Tissue 6005-N	6.7
Liver Margin (ODO4310)	23.8	Normal Bladder	54.7
Melanoma Mets to Lung (OD04321)	16.7	Bladder Cancer 1023	7.6
Lung Margin (OD04321)	20.6	Bladder Cancer A302173	23.2
Normal Kidney	16.6	Bladder Cancer (OD04718-01)	100.0
Kidney Ca, Nuclear grade 2 (OD04338)	15.8	Bladder Normal Adjacent (OD04718-03)	30.4
Kidney Margin (OD04338)	11.5	Normal Ovary	2.3
Kidney Ca Nuclear grade 1/2 (OD04339)	10.2	Ovarian Cancer 064008	30.6
Kidney Margin (OD04339)	15.9	Ovarian Cancer (OD04768-07)	37.6
Kidney Ca, Clear cell type (OD04340)	28.7	Ovary Margin (OD04768-08)	14.8
Kidney Margin (OD04340)	15.5	Normal Stomach	17.9
Kidney Ca, Nuclear grade 3 (OD04348)	12.3	Gastric Cancer 9060358	10.0
Kidney Margin (OD04348)	16.0	Stomach Margin 9060359	23.0
Kidney Cancer (OD04622-01)	12.5	Gastric Cancer 9060395	35.8
Kidney Margin (OD04622-03)	3.6	Stomach Margin 9060394	27.5
Kidney Cancer (OD04450-01)	9.5	Gastric Cancer 9060397	66.0
Kidney Margin (OD04450-03)	13.0	Stomach Margin 9060396	18.0
Kidney Cancer 8120607	3.1	Gastric Cancer 064005	62.0

Table BBE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2900, Run 159996820	Tissue Name	Rel. Exp.(%) Ag2900, Run 159996820
Secondary Th1 act	68.3	HUVEC IL-1beta	10.5
Secondary Th2 act	67.8	HUVEC IFN gamma	15.5

Secondary Tr1 act	73.2	HUVEC TNF alpha + IFN gamma	16.0
Secondary Th1 rest	8.2	HUVEC TNF alpha + IL4	18.8
Secondary Th2 rest	18.6	HUVEC IL-11	13.1
Secondary Tr1 rest	10.1	Lung Microvascular EC none	18.8
Primary Th1 act	75.8	Lung Microvascular EC TNFalpha + IL-1beta	25.2
Primary Th2 act	60.7	Microvascular Dermal EC none	30.8
Primary Tr1 act	74.7	Microsvascular Dermal EC TNFalpha + IL-1beta	27.9
Primary Th1 rest	42.0	Bronchial epithelium TNFalpha + IL1beta	16.8
Primary Th2 rest	29.3	Small airway epithelium none	11.7
Primary Tr1 rest	16.3	Small airway epithelium TNFalpha + IL-1beta	79.0
CD45RA CD4 lymphocyte act	34.4	Coronary artery SMC rest	19.2
CD45RO CD4 lymphocyte act	53.6	Coronary artery SMC TNFalpha + IL-1beta	7.6
CD8 lymphocyte act	35.8	Astrocytes rest	8.6
Secondary CD8 lymphocyte rest	39.0	Astrocytes TNFalpha + IL-1beta	7.2
Secondary CD8 lymphocyte act	33.9	KU-812 (Basophil) rest	7.8
CD4 lymphocyte none	6.2	KU-812 (Basophil) PMA/ionomycin	40.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	11.7	CCD1106 (Keratinocytes) none	30.4
LAK cells rest	25.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	13.6
LAK cells IL-2	23.8	Liver cirrhosis	3.9
LAK cells IL-2+IL-12	22.5	Lupus kidney	1.3
LAK cells IL-2+IFN gamma	36.3	NCI-H292 none	87.1
LAK cells IL-2+ IL-18	29.7	NCI-H292 IL-4	91.4
LAK cells PMA/ionomycin	47.0	NCI-H292 IL-9	92.0
NK Cells IL-2 rest	13.7	NCI-H292 IL-13	39.8
Two Way MLR 3 day	15.7	NCI-H292 IFN gamma	47.0
Two Way MLR 5 day	15.6	HPAEC none	18.0
Two Way MLR 7 day	12.6	HPAEC TNF alpha + IL-1 beta	26.1

PBMC rest	7.5	Lung fibroblast none	22.2
PBMC PWM	100.0	Lung fibroblast TNF alpha + IL-1 beta	16.8
PBMC PHA-L	42.6	Lung fibroblast IL-4	62.4
Ramos (B cell) none	17.0	Lung fibroblast IL-9	39.5
Ramos (B cell) ionomycin	54.0	Lung fibroblast IL-13	35.4
B lymphocytes PWM	73.7	Lung fibroblast IFN gamma	70.2
B lymphocytes CD40L and IL-4	12.9	Dermal fibroblast CCD1070 rest	49.3
EOL-1 dbcAMP	9.5	Dermal fibroblast CCD1070 TNF alpha	92.7
EOL-1 dbcAMP PMA/ionomycin	22.8	Dermal fibroblast CCD1070 IL-1 beta	29.5
Dendritic cells none	13.7	Dermal fibroblast IFN gamma	27.7
Dendritic cells LPS	15.0	Dermal fibroblast IL-4	41.5
Dendritic cells anti-CD40	19.1	IBD Colitis 2	1.1
Monocytes rest	10.7	IBD Crohn's	2.6
Monocytes LPS	4.3	Colon	14.8
Macrophages rest	26.1	Lung	21.5
Macrophages LPS	15.3	Thymus	25.0
HUVEC none	30.8	Kidney	37.4
HUVEC starved	34.6		

CNS_neurodegeneration_v1.0 Summary: Ag2900 This panel confirms the expression of this gene at low levels in the brain in an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.3D for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

Panel 1.3D Summary: Ag2900 The CG56633-01 gene is expressed at moderate levels in the cancer cell lines in this panel, with highest expression in a breast cancer cell line (CT=27). Expression of this gene could potentially be used as a diagnostic marker of cell proliferation and hence as a diagnostic marker for cancer.

This gene also has moderate levels of expression in adipose, liver, heart, skeletal muscle, adrenal, pituitary, thyroid and pancreas. Therefore, therapeutic modulation of this

gene product may be a treatment for endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

Overall, this gene, a translation initiation factor homolog, exhibits brain-preferential expression, particularly in the hippocampus, a structure critical for learning and memory. The processes of learning and memory are subject to regulation by mechanisms of translational and transcriptional control, including the regulation elongation factor phosphorylation by the memory-mediating NMDA receptor. The hippocampus-preferential expression of this gene suggests that it plays a role in translationally-mediated learning and memory processes. Therefore, agents that modulate the activity and function of this gene product may have utility in treating CNS disorders involving memory deficits, including Alzheimer's disease and aging.

References:

Scheetz AJ, Nairn AC, Constantine-Paton M. N-methyl-D-aspartate receptor activation and visual activity induce elongation factor-2 phosphorylation in amphibian tecta: a role for N-methyl-D-aspartate receptors in controlling protein synthesis. *Proc Natl Acad Sci U S A* 1997 Dec 23;94(26):14770-5

N-methyl-D-aspartate receptor (NMDAR) activation has been implicated in forms of synaptic plasticity involving long-term changes in neuronal structure, function, or protein expression. Transcriptional alterations have been correlated with NMDAR-mediated synaptic plasticity, but the problem of rapidly targeting new proteins to particular synapses is unsolved. One potential solution is synapse-specific protein translation, which is suggested by dendritic localization of numerous transcripts and subsynaptic polyribosomes. We report here a mechanism by which NMDAR activation at synapses may control this protein synthetic machinery. In intact tadpole tecta, NMDAR activation leads to phosphorylation of a subset of proteins, one of which we now identify as the eukaryotic translation elongation factor 2 (eEF2). Phosphorylation of eEF2 halts protein synthesis and may prepare cells to translate a new set of mRNAs. We show that NMDAR activation-induced eEF2 phosphorylation is widespread in tadpole tecta. In contrast, in adult tecta, where synaptic plasticity is reduced, this phosphorylation is restricted to short dendritic regions that process binocular information. Biochemical and anatomical evidence shows that this NMDAR activation-induced eEF2 phosphorylation is localized to subsynaptic sites. Moreover, eEF2 phosphorylation is induced by visual stimulation, and NMDAR blockade before stimulation eliminates this effect. Thus, NMDAR activation, which is known to mediate synaptic changes in the developing frog, could produce local postsynaptic alterations in protein synthesis by inducing eEF2 phosphorylation.

Panel 2D Summary: Ag2900 The CG56633-01 gene is expressed at increased levels in colon, breast and bladder cancers compared to the normal adjacent tissue samples. Therefore, expression of this gene could be of use as a marker for these cancers.

Panel 4D Summary: Ag2900 The CG56633-01 gene is expressed in a number of preparations of activated T lymphocytes at levels greater than in resting T cells. Therefore, small molecule antagonists of the CG56633-01 gene product may reduce T cell activation and thus reduce or eliminate the symptoms in patients with autoimmune and inflammatory diseases, such as Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

NOV60a and NOV60b

Expression of gene NOV60a and variant NOV60b was assessed using the primer-probe sets Ag041b and Ag41, described in Tables BCA and BCB. Results of the RTQ-PCR runs are shown in Tables BCC, BCD, BCE, BCF, BCG, BCH and BCI.

Table BCA. Probe Name Ag041b

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtagtaggtgcgctgggtcat-3'	21	486	1231
Probe	TET-5'-accatagccgggcagcgcagtg-3'-TAMRA	21	455	1232
Reverse	5'-caacggagacaactgcttcaac-3'	22	431	1233

Table BCB. Probe Name Ag41

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gtaggtgcgctgggtcatg-3'	19	485	1234
Probe	TET-5'-ccatgcgctgcccgggtatg-3'-TAMRA	20	454	1235
Reverse	5'-cctacaacggagacaactgcttc-3'	23	427	1236

Table BCC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag041b, Run 206231412	Tissue Name	Rel. Exp.(%) Ag041b, Run 206231412
AD 1 Hippo	20.7	Control (Path) 3 Temporal Ctx	9.7
AD 2 Hippo	44.1	Control (Path) 4 Temporal Ctx	70.2
AD 3 Hippo	9.5	AD 1 Occipital Ctx	17.9
AD 4 Hippo	17.4	AD 2 Occipital Ctx	0.0

		(Missing)	
AD 5 Hippo	97.9	AD 3 Occipital Ctx	5.8
AD 6 Hippo	65.1	AD 4 Occipital Ctx	39.8
Control 2 Hippo	66.0	AD 5 Occipital Ctx	76.8
Control 4 Hippo	18.4	AD 6 Occipital Ctx	20.0
Control (Path) 3 Hippo	11.3	Control 1 Occipital Ctx	5.3
AD 1 Temporal Ctx	11.3	Control 2 Occipital Ctx	82.4
AD 2 Temporal Ctx	54.3	Control 3 Occipital Ctx	36.9
AD 3 Temporal Ctx	9.5	Control 4 Occipital Ctx	12.1
AD 4 Temporal Ctx	39.8	Control (Path) 1 Occipital Ctx	95.3
AD 5 Inf Temporal Ctx	80.7	Control (Path) 2 Occipital Ctx	33.4
AD 5 Sup Temporal Ctx	57.8	Control (Path) 3 Occipital Ctx	4.4
AD 6 Inf Temporal Ctx	52.1	Control (Path) 4 Occipital Ctx	53.6
AD 6 Sup Temporal Ctx	54.3	Control 1 Parietal Ctx	9.3
Control 1 Temporal Ctx	13.8	Control 2 Parietal Ctx	49.0
Control 2 Temporal Ctx	66.0	Control 3 Parietal Ctx	34.6
Control 3 Temporal Ctx	36.3	Control (Path) 1 Parietal Ctx	100.0
Control 3 Temporal Ctx	19.8	Control (Path) 2 Parietal Ctx	44.4
Control (Path) 1 Temporal Ctx	85.3	Control (Path) 3 Parietal Ctx	6.9
Control (Path) 2 Temporal Ctx	73.7	Control (Path) 4 Parietal Ctx	72.2

Table BCD. Panel 1

Tissue Name	Rel. Exp.(%) Ag41, Run 97804013	Rel. Exp.(%) Ag41, Run 97807227	Tissue Name	Rel. Exp.(%) Ag41, Run 97804013	Rel. Exp.(%) Ag41, Run 97807227
Endothelial cells	0.0	0.2	Renal ca. 786-0	0.0	0.4
Endothelial cells (treated)	0.0	0.3	Renal ca. A498	11.6	32.8
Pancreas	0.1	4.2	Renal ca. RXF	3.6	1.0

			393		
Pancreatic ca. CAPAN 2	0.0	0.5	Renal ca. ACHN	0.0	1.1
Adrenal gland	1.2	15.3	Renal ca. UO- 31	0.0	0.9
Thyroid	0.0	4.4	Renal ca. TK- 10	0.0	0.2
Salivary gland	0.1	6.5	Liver	0.9	8.3
Pituitary gland	0.2	6.0	Liver (fetal)	0.1	3.3
Brain (fetal)	0.0	3.7	Liver ca. (hepatoblast) HepG2	0.2	5.6
Brain (whole)	0.0	35.4	Lung	0.0	1.8
Brain (amygdala)	6.4	60.7	Lung (fetal)	0.0	1.1
Brain (cerebellum)	4.2	25.5	Lung ca. (small cell) LX-1	0.0	1.0
Brain (hippocampus)	6.1	54.0	Lung ca. (small cell) NCI-H69	0.0	0.7
Brain (substantia nigra)	3.4	35.6	Lung ca. (s.cell var.) SHP-77	0.0	0.2
Brain (thalamus)	11.1	100.0	Lung ca. (large cell)NCI-H460	15.9	2.7
Brain (hypothalamus)	0.2	6.3	Lung ca. (non- sm. cell) A549	0.0	1.4
Spinal cord	4.1	14.0	Lung ca. (non- s.cell) NCI-H23	5.7	14.7
glio/astro U87- MG	0.0	0.0	Lung ca. (non- s.cell) HOP-62	0.0	0.4
glio/astro U-118- MG	0.0	0.2	Lung ca. (non- s.cl) NCI-H522	0.2	6.3
astrocytoma SW1783	0.0	0.4	Lung ca. (squam.) SW 900	0.3	6.9
neuro*; met SK- N-AS	0.0	0.3	Lung ca. (squam.) NCI- H596	0.0	0.4
astrocytoma SF- 539	0.0	0.8	Mammary gland	100.0	10.5
astrocytoma SNB- 75	0.3	4.7	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SNB-19	0.3	2.3	Breast ca.* (pl.ef) MDA- MB-231	0.0	0.2
glioma U251	7.6	0.5	Breast ca.* (pl. ef) T47D	0.8	7.4
glioma SF-295	0.0	0.8	Breast ca. BT-	0.0	0.1

			549		
Heart	6.8	49.0	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	0.5	6.0	Ovary	0.9	13.0
Bone marrow	0.0	1.0	Ovarian ca. OVCAR-3	0.0	1.0
Thymus	0.0	0.9	Ovarian ca. OVCAR-4	0.0	0.4
Spleen	0.0	2.5	Ovarian ca. OVCAR-5	0.4	7.1
Lymph node	0.0	1.1	Ovarian ca. OVCAR-8	0.0	0.2
Colon (ascending)	0.0	1.7	Ovarian ca. IGROV-1	0.0	0.5
Stomach	4.6	3.2	Ovarian ca. (ascites) SK- OV-3	3.8	0.7
Small intestine	0.1	3.4	Uterus	0.7	11.7
Colon ca. SW480	0.0	0.3	Placenta	0.9	9.9
Colon ca.* SW620 (SW480 met)	0.0	0.6	Prostate	0.3	7.7
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met) PC- 3	0.0	0.1
Colon ca. HCT- 116	0.0	0.8	Testis	0.1	3.0
Colon ca. CaCo-2	0.0	0.5	Melanoma Hs688(A).T	0.0	0.9
Colon ca. HCT-15	0.0	0.4	Melanoma* (met) Hs688(B).T	0.0	2.3
Colon ca. HCC- 2998	0.0	0.0	Melanoma UACC-62	0.0	0.3
Gastric ca. * (liver met) NCI-N87	0.0	0.3	Melanoma M14	0.0	0.1
Bladder	0.0	2.5	Melanoma LOX IMVI	0.0	0.0
Trachea	0.0	2.0	Melanoma* (met) SK-MEL- 5	0.0	0.1
Kidney	1.0	11.6	Melanoma SK- MEL-28	0.1	1.5
Kidney (fetal)	0.0	3.4			

Table BCE. Panel 1.1

Tissue Name	Rel. Exp.(%) Ag041b, Run 109666937	Tissue Name	Rel. Exp.(%) Ag041b, Run 109666937
Adrenal gland	11.5	Renal ca. UO-31	1.3
Bladder	2.4	Renal ca. RXF 393	0.4
Brain (amygdala)	25.2	Liver	9.9
Brain (cerebellum)	57.4	Liver (fetal)	2.9
Brain (hippocampus)	77.4	Liver ca. (hepatoblast) HepG2	2.7
Brain (substantia nigra)	57.4	Lung	1.6
Brain (thalamus)	47.6	Lung (fetal)	3.3
Cerebral Cortex	100.0	Lung ca. (non-s.cell) HOP-62	0.3
Brain (fetal)	5.5	Lung ca. (large cell)NCI-H460	1.4
Brain (whole)	71.2	Lung ca. (non-s.cell) NCI-H23	6.2
glio/astro U-118-MG	0.2	Lung ca. (non-s.cl) NCI-H522	9.2
astrocytoma SF-539	1.1	Lung ca. (non-sm. cell) A549	1.7
astrocytoma SNB-75	5.7	Lung ca. (s.cell var.) SHP-77	0.2
astrocytoma SW1783	0.6	Lung ca. (small cell) LX-1	1.4
glioma U251	0.2	Lung ca. (small cell) NCI-H69	0.6
glioma SF-295	1.2	Lung ca. (squam.) SW 900	5.9
glioma SNB-19	2.8	Lung ca. (squam.) NCI-H596	0.5
glio/astro U87-MG	0.0	Lymph node	1.4
neuro*; met SK-N-AS	0.2	Spleen	2.5
Mammary gland	6.3	Thymus	0.5
Breast ca. BT-549	0.0	Ovary	6.7
Breast ca. MDA-N	0.0	Ovarian ca. IGROV- 1	0.5
Breast ca.* (pl.ef) T47D	8.0	Ovarian ca. OVCAR-3	1.1
Breast ca.* (pl.ef) MCF-7	0.0	Ovarian ca. OVCAR-4	0.2
Breast ca.* (pl.ef) MDA-MB-231	0.3	Ovarian ca. OVCAR-5	9.8
Small intestine	6.7	Ovarian ca.	0.0

		OVCAR-8	
Colorectal	0.5	Ovarian ca.* (ascites) SK-OV-3	0.7
Colon ca. HT29	0.0	Pancreas	11.5
Colon ca. CaCo-2	0.9	Pancreatic ca. CAPAN 2	0.4
Colon ca. HCT-15	0.1	Pituitary gland	6.6
Colon ca. HCT-116	0.4	Placenta	9.5
Colon ca. HCC-2998	0.0	Prostate	7.0
Colon ca. SW480	0.1	Prostate ca.* (bone met) PC-3	0.0
Colon ca.* SW620 (SW480 met)	1.3	Salivary gland	6.2
Stomach	8.0	Trachea	2.3
Gastric ca. (liver met) NCI-N87	0.3	Spinal cord	14.3
Heart	92.7	Testis	3.1
Skeletal muscle (Fetal)	4.9	Thyroid	6.1
Skeletal muscle	16.8	Uterus	4.7
Endothelial cells	1.0	Melanoma M14	0.0
Heart (Fetal)	33.9	Melanoma LOX IMVI	0.0
Kidney	16.8	Melanoma UACC- 62	0.1
Kidney (fetal)	3.1	Melanoma SK-MEL- 28	0.7
Renal ca. 786-0	0.2	Melanoma* (met) SK-MEL-5	0.0
Renal ca. A498	30.8	Melanoma Hs688(A).T	1.1
Renal ca. ACHN	1.2	Melanoma* (met) Hs688(B).T	1.5
Renal ca. TK-10	0.2		

Table BCF. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag041b, Run 150010102	Tissue Name	Rel. Exp.(%) Ag041b, Run 150010102
Liver adenocarcinoma	0.0	Kidney (fetal)	0.5
Pancreas	0.6	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.1	Renal ca. A498	11.0
Adrenal gland	2.0	Renal ca. RXF	0.1

		393	
Thyroid	1.0	Renal ca. ACHN	0.2
Salivary gland	0.6	Renal ca. UO-31	0.3
Pituitary gland	0.8	Renal ca. TK-10	0.0
Brain (fetal)	0.5	Liver	1.0
Brain (whole)	25.0	Liver (fetal)	0.8
Brain (amygdala)	25.3	Liver ca. (hepatoblast) HepG2	0.6
Brain (cerebellum)	2.7	Lung	0.6
Brain (hippocampus)	100.0	Lung (fetal)	0.9
Brain (substantia nigra)	3.5	Lung ca. (small cell) LX-1	0.1
Brain (thalamus)	20.7	Lung ca. (small cell) NCI-H69	0.2
Cerebral Cortex	63.3	Lung ca. (s.cell var.) SHP-77	0.1
Spinal cord	3.6	Lung ca. (large cell) NCI-H460	0.1
MG glio/astro U87-	0.0	Lung ca. (non- sm. cell) A549	0.2
MG glio/astro U-118-	0.1	Lung ca. (non- s.cell) NCI-H23	3.4
SW1783 astrocytoma	0.1	Lung ca. (non- s.cell) HOP-62	0.1
N-AS neuro*; met SK-	0.0	Lung ca. (non- s.cl) NCI-H522	1.2
539 astrocytoma SF-	0.1	Lung ca. (squam.) SW 900	0.5
SNB-75 astrocytoma	0.8	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.4	Mammary gland	1.2
glioma U251	0.1	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.4	Breast ca.* (pl.ef) MDA-MB-231	0.1
Heart (fetal)	21.9	Breast ca.* (pl.ef) T47D	1.0
Heart	3.7	Breast ca. BT- 549	0.0
Skeletal muscle (fetal)	10.7	Breast ca. MDA- N	0.0
Skeletal muscle	0.2	Ovary	8.6
Bone marrow	0.2	Ovarian ca.	0.1

		OVCAR-3	
Thymus	0.1	Ovarian ca. OVCAR-4	0.1
Spleen	1.7	Ovarian ca. OVCAR-5	0.8
Lymph node	0.4	Ovarian ca. OVCAR-8	0.0
Colorectal	1.7	Ovarian ca. IGROV-1	0.1
Stomach	1.5	Ovarian ca.* (ascites) SK-OV-3	0.1
Small intestine	1.6	Uterus	1.9
Colon ca. SW480	0.2	Placenta	1.7
Colon ca.* SW620(SW480 met)	0.2	Prostate	0.9
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT- 116	0.1	Testis	0.8
Colon ca. CaCo- 2	0.1	Melanoma Hs688(A).T	1.3
Colon ca. tissue(ODO3866)	0.3	Melanoma* (met) Hs688(B).T	0.7
Colon ca. HCC- 2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.6	Melanoma LOX IMVI	0.0
Trachea	0.7	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.7	Adipose	1.6

Table BCG. Panel 2D

Tissue Name	Rel. Exp.(%) Ag041b, Run 157096248	Rel. Exp.(%) Ag41, Run 157938256	Tissue Name	Rel. Exp.(%) Ag041b, Run 157096248	Rel. Exp.(%) Ag41, Run 157938256
Normal Colon	15.7	21.5	Kidney Margin 8120608	9.2	12.9
CC Well to Mod Diff (ODO3866)	4.4	2.9	Kidney Cancer 8120613	4.9	7.2
CC Margin (ODO3866)	6.3	7.3	Kidney Margin	14.4	16.6

			8120614		
CC Gr.2 rectosigmoid (ODO3868)	3.3	2.3	Kidney Cancer 9010320	63.7	69.3
CC Margin (ODO3868)	8.2	8.6	Kidney Margin 9010321	33.0	42.3
CC Mod Diff (ODO3920)	4.9	6.9	Normal Uterus	8.5	14.9
CC Margin (ODO3920)	8.8	11.7	Uterus Cancer 064011	16.2	14.3
CC Gr.2 ascend colon (ODO3921)	2.4	1.9	Normal Thyroid	5.7	8.4
CC Margin (ODO3921)	3.4	4.1	Thyroid Cancer 064010	8.4	12.1
CC from Partial Hepatectomy (ODO4309) Mets	2.1	7.1	Thyroid Cancer A302152	8.5	9.9
Liver Margin (ODO4309)	15.1	24.3	Thyroid Margin A302153	9.2	12.6
Colon mets to lung (OD04451-01)	0.8	1.2	Normal Breast	24.3	42.3
Lung Margin (OD04451-02)	2.3	3.1	Breast Cancer (OD04566)	1.8	2.1
Normal Prostate 6546-1	9.7	9.0	Breast Cancer (OD04590-01)	5.2	3.9
Prostate Cancer (OD04410)	13.6	12.4	Breast Cancer Mets (OD04590-03)	8.5	9.9
Prostate Margin (OD04410)	18.3	19.3	Breast Cancer Metastasis (OD04655-05)	8.8	11.7
Prostate Cancer (OD04720-01)	19.6	28.1	Breast Cancer 064006	10.3	12.9
Prostate Margin (OD04720-02)	33.4	36.9	Breast Cancer 1024	15.4	16.2
Normal Lung 061010	7.2	10.9	Breast Cancer 9100266	3.8	5.7
Lung Met to Muscle (ODO4286)	0.7	1.2	Breast Margin 9100265	4.0	6.3
Muscle Margin (ODO4286)	7.9	9.7	Breast Cancer A209073	4.4	7.2
Lung Malignant	3.0	3.7	Breast Margin	8.2	8.4

Kidney Ca, Nuclear grade 3 (OD04348)	3.1	3.3	Gastric Cancer 9060358	3.5	3.9
Kidney Margin (OD04348)	8.4	12.7	Stomach Margin 9060359	3.0	3.7
Kidney Cancer (OD04622-01)	3.6	4.0	Gastric Cancer 9060395	6.7	6.3
Kidney Margin (OD04622-03)	4.5	4.7	Stomach Margin 9060394	6.1	7.8
Kidney Cancer (OD04450-01)	1.6	0.7	Gastric Cancer 9060397	9.2	13.8
Kidney Margin (OD04450-03)	9.9	13.9	Stomach Margin 9060396	3.3	3.0
Kidney Cancer 8120607	2.4	1.8	Gastric Cancer 064005	4.7	5.0

Table BCH. Panel 3D

Tissue Name	Rel. Exp.(%) Ag041b, Run 156897045	Rel. Exp.(%) Ag41, Run 157938257	Tissue Name	Rel. Exp.(%) Ag041b, Run 156897045	Rel. Exp.(%) Ag41, Run 157938257
Daoy- Medulloblastoma	0.4	0.2	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0	0.0
TE671- Medulloblastoma	0.3	0.2	ES-2- Ovarian clear cell carcinoma	0.0	0.0
D283 Med- Medulloblastoma	0.1	0.3	Ramos- Stimulated with PMA/ionomycin 6h	0.0	0.0
PFSK-1- Primitive Neuroectodermal	0.1	0.5	Ramos- Stimulated with PMA/ionomycin 14h	0.0	0.0
XF-498- CNS	1.8	2.7	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0	0.3
SNB-78- Glioma	0.0	0.0	Raji- Burkitt's lymphoma	0.0	0.0
SF-268- Glioblastoma	0.4	0.7	Daudi- Burkitt's lymphoma	0.3	0.0
T98G- Glioblastoma	1.9	2.0	U266- B-cell plasmacytoma	0.0	0.0

SK-N-SH- Neuroblastoma (metastasis)	2.3	2.9	CA46- Burkitt's lymphoma	0.1	0.0
SF-295- Glioblastoma	0.5	1.1	RL- non-Hodgkin's B-cell lymphoma	0.0	0.0
Cerebellum	30.8	55.9	JM1- pre-B-cell lymphoma	0.0	0.0
Cerebellum	32.3	60.3	Jurkat- T cell leukemia	0.0	0.0
NCI-H292- Mucoepidermoid lung carcinoma	2.1	1.7	TF-1- Erythroleukemia	0.0	0.1
DMS-114- Small cell lung cancer	6.0	6.4	HUT 78- T-cell lymphoma	0.0	0.0
DMS-79- Small cell lung cancer	100.0	100.0	U937- Histiocytic lymphoma	0.2	0.0
NCI-H146- Small cell lung cancer	0.0	0.3	KU-812- Myelogenous leukemia	0.8	0.8
NCI-H526- Small cell lung cancer	2.2	3.8	769-P- Clear cell renal carcinoma	0.5	0.1
NCI-N417- Small cell lung cancer	2.7	4.5	Caki-2- Clear cell renal carcinoma	1.4	1.4
NCI-H82- Small cell lung cancer	0.1	0.3	SW 839- Clear cell renal carcinoma	0.0	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	0.0	G401- Wilms' tumor	0.1	0.0
NCI-H1155- Large cell lung cancer	2.1	2.6	Hs766T- Pancreatic carcinoma (LN metastasis)	1.1	1.2
NCI-H1299- Large cell lung cancer	0.1	0.3	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	7.5	13.0
NCI-H727- Lung carcinoid	0.0	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	1.9	2.9
NCI-UMC-11- Lung carcinoid	0.1	0.0	BxPC-3- Pancreatic adenocarcinoma	5.0	7.0
LX-1- Small cell lung cancer	1.3	1.4	HPAC- Pancreatic adenocarcinoma	0.4	0.4
Colo-205- Colon cancer	0.0	0.0	MIA PaCa-2- Pancreatic carcinoma	3.6	4.0
KM12- Colon cancer	0.0	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	1.6	2.0

KM20L2- Colon cancer	0.0	0.1	PANC-1- Pancreatic epithelioid ductal carcinoma	20.3	22.1
NCI-H716- Colon cancer	0.1	0.0	T24- Bladder carcinoma (transitional cell)	0.4	0.4
SW-48- Colon adenocarcinoma	0.1	0.0	5637- Bladder carcinoma	1.2	0.7
SW1116- Colon adenocarcinoma	0.0	0.0	HT-1197- Bladder carcinoma	0.0	0.2
LS 174T- Colon adenocarcinoma	0.3	0.1	UM-UC-3- Bladder carcinoma (transitional cell)	0.1	0.0
SW-948- Colon adenocarcinoma	0.0	0.0	A204- Rhabdomyosarcoma	2.3	3.5
SW-480- Colon adenocarcinoma	0.0	0.0	HT-1080- Fibrosarcoma	0.2	0.4
NCI-SNU-5- Gastric carcinoma	0.5	0.6	MG-63- Osteosarcoma	9.3	20.9
KATO III- Gastric carcinoma	0.4	0.4	SK-LMS-1- Leiomyosarcoma (vulva)	0.1	0.4
NCI-SNU-16- Gastric carcinoma	0.1	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0	0.0
NCI-SNU-1- Gastric carcinoma	0.2	0.0	A431- Epidermoid carcinoma	0.4	0.3
RF-1- Gastric adenocarcinoma	0.3	0.0	WM266-4- Melanoma	0.6	0.5
RF-48- Gastric adenocarcinoma	0.0	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0	0.0
MKN-45- Gastric carcinoma	0.7	0.4	MDA-MB-468- Breast adenocarcinoma	26.4	30.4
NCI-N87- Gastric carcinoma	0.2	0.1	SCC-4- Squamous cell carcinoma of tongue	0.2	0.0
OVCAR-5- Ovarian carcinoma	1.7	1.7	SCC-9- Squamous cell carcinoma of tongue	0.2	0.2
RL95-2- Uterine carcinoma	5.2	9.1	SCC-15- Squamous cell carcinoma of tongue	0.6	0.3
HelaS3- Cervical adenocarcinoma	0.0	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0	0.1

Table BCI. Panel 4D

Tissue Name	Rel. Exp.(%) Ag041b, Run 146087302	Tissue Name	Rel. Exp.(%) Ag041b, Run 146087302
Secondary Th1 act	2.1	HUVEC IL-1beta	1.6
Secondary Th2 act	0.9	HUVEC IFN gamma	11.0
Secondary Tr1 act	0.9	HUVEC TNF alpha + IFN gamma	5.5
Secondary Th1 rest	1.2	HUVEC TNF alpha + IL4	2.9
Secondary Th2 rest	2.7	HUVEC IL-11	5.8
Secondary Tr1 rest	3.0	Lung Microvascular EC none	2.7
Primary Th1 act	0.9	Lung Microvascular EC TNFalpha + IL-1beta	0.8
Primary Th2 act	1.6	Microvascular Dermal EC none	2.4
Primary Tr1 act	2.6	Microvascular Dermal EC TNFalpha + IL-1beta	2.0
Primary Th1 rest	2.7	Bronchial epithelium TNFalpha + IL1beta	7.5
Primary Th2 rest	2.7	Small airway epithelium none	2.3
Primary Tr1 rest	0.8	Small airway epithelium TNFalpha + IL-1beta	6.2
CD45RA CD4 lymphocyte act	2.4	Coronary artery SMC rest	6.6
CD45RO CD4 lymphocyte act	3.9	Coronary artery SMC TNFalpha + IL-1beta	10.9
CD8 lymphocyte act	0.7	Astrocytes rest	10.2
Secondary CD8 lymphocyte rest	2.0	Astrocytes TNFalpha + IL-1beta	13.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	2.3
CD4 lymphocyte none	1.9	KU-812 (Basophil) PMA/ionomycin	2.5
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	2.2
LAK cells rest	3.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.8
LAK cells IL-2	1.3	Liver cirrhosis	3.9
LAK cells IL-2+IL-12	1.3	Lupus kidney	5.6
LAK cells IL-2+IFN gamma	2.2	NCI-H292 none	5.3

LAK cells IL-2+ IL-18	0.8	NCI-H292 IL-4	6.2
LAK cells PMA/ionomycin	1.3	NCI-H292 IL-9	4.9
NK Cells IL-2 rest	5.1	NCI-H292 IL-13	9.2
Two Way MLR 3 day	1.8	NCI-H292 IFN gamma	7.1
Two Way MLR 5 day	0.2	HPAEC none	3.6
Two Way MLR 7 day	0.4	HPAEC TNF alpha + IL-1 beta	1.1
PBMC rest	0.9	Lung fibroblast none	74.2
PBMC PWM	1.6	Lung fibroblast TNF alpha + IL-1 beta	33.4
PBMC PHA-L	0.3	Lung fibroblast IL-4	100.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	62.9
Ramos (B cell) ionomycin	1.3	Lung fibroblast IL-13	51.4
B lymphocytes PWM	0.8	Lung fibroblast IFN gamma	84.1
B lymphocytes CD40L and IL-4	0.4	Dermal fibroblast CCD1070 rest	6.7
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	6.7
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	15.3
Dendritic cells none	12.2	Dermal fibroblast IFN gamma	40.6
Dendritic cells LPS	7.7	Dermal fibroblast IL-4	83.5
Dendritic cells anti- CD40	12.2	IBD Colitis 2	3.9
Monocytes rest	6.3	IBD Crohn's	4.5
Monocytes LPS	0.7	Colon	20.4
Macrophages rest	6.7	Lung	11.5
Macrophages LPS	0.9	Thymus	41.5
HUVEC none	5.3	Kidney	4.1
HUVEC starved	6.3		

CNS_neurodegeneration_v1.0 Summary: Ag041b This panel does not show differential expression of the NOV60a gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1 Summary: Ag41 Two experiments with the same probe and primer set produce results that are in reasonable agreement, with highest expression of the NOV60a gene in the mammary gland and the brain. Overall, this gene appears to express in normal tissues at

higher levels than in cancer cell lines. This gene encodes a lynx1 homolog. Lynx1 is an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. Activation of nicotinic receptors is associated with positive effect on schizophrenia and alzheimer's disease. Therefore, agents that block Ag41 action in the CNS are likely to have utility in the treatment of these, and related, disorders.

This gene also has high levels of expression in pancreas, adrenal, thyroid, pituitary, heart, skeletal muscle and liver. Therefore, therapeutic modulation of this gene product may be a treatment for endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes. Please note that two additional experiments with the same probe and primer set show low/undetectable levels of expression (CTs>35). (Data not shown.) The results indicate that there is a possibility of a probe failure.

References:

Miwa JM, Ibanez-Tallon I, Crabtree GW, Sanchez R, Sali A, Role LW, Heintz N. lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. Neuron 1999 May;23(1):105-14

Elapid snake venom neurotoxins exert their effects through high-affinity interactions with specific neurotransmitter receptors. A novel murine gene, lynx1, is highly expressed in the brain and contains the cysteine-rich motif characteristic of this class of neurotoxins. Primary sequence and gene structure analyses reveal an evolutionary relationship between lynx1 and the Ly-6/neurotoxin gene family. lynx1 is expressed in large projection neurons in the hippocampus, cortex, and cerebellum. In cerebellar neurons, lynx1 protein is localized to a specific subdomain including the soma and proximal dendrites. lynx1 binding to brain sections correlates with the distribution of nAChRs, and application of lynx1 to Xenopus oocytes expressing nAChRs results in an increase in acetylcholine-evoked macroscopic currents. These results identify lynx1 as a novel protein modulator for nAChRs in vitro, which could have important implications in the regulation of cholinergic function in vivo.

Panel 1.1 Summary: Ag041b The NOV60a gene is expressed in most cell lines and normal tissues with a significantly higher level of expression in normal brain and heart compared to cancer cell lines on this panel. The results in this panel are consistent with expression in Panel 1. Please see Panel 1 for further discussion of utility of this gene in metabolic and cns diseases and cancer.

Panel 1.3D Summary: Ag041b Highest expression of the NOV60a gene is seen in the brain. Overall, this gene is expressed in most cell lines and normal tissues with a significantly higher level of expression in heart in addition to brain when compared to cancer cell lines on

this panel. Please see Panel 1 for discussion of utility of this gene in the central nervous system.

Among metabolic tissues, this gene has a low level of expression in adipose, adult and fetal liver, adrenal, pituitary, fetal skeletal muscle, fetal and adult heart, thyroid and pancreas.

5 Therefore, modulation of this gene product may be a treatment for endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes. In addition, this gene differentially expressed in fetal (CT value = 29) vs adult skeletal muscle (CT value = 35) and may be useful for the identification of the two sources of this tissue.

Panel 2D Summary: Ag041b/Ag41 The expression of the NOV60a gene was assessed
10 in two independent runs on this panel with good concordance between runs. This protein is a good diagnostic marker and target in ovarian, renal and liver cancer as the cancer expressed this gene at a higher level than the normal adjacent tissue.

Panel 3D Summary: Ag041b/Ag41 Two experiments show expression of the
15 NOV60a gene in cell lines derived from brain, lung, ovarian, renal, pancreatic, breast and osteosarcoma. Therefore, expression of this gene could be used as a diagnostic marker for the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of brain, lung, ovarian, renal, pancreatic, breast and osteosarcoma cancers.

Panel 4D Summary: Ag041b The NOV60a gene, a lynx1 homolog is expressed at
20 moderate levels in untreated lung fibroblasts, lung fibroblasts activated with IL-4, IL-9 or IFN gamma, and dermal fibroblasts activated with IL-4 (CTs=30). Therefore, small molecules or therapeutic antibodies that antagonize the function of the NOV60a gene product may be useful to reduce or eliminate the symptoms in patients with chronic obstructive pulmonary disease, asthma, emphysema, or psoriasis.

25 NOV61: Adlican-like

Expression of gene NOV61 was assessed using the primer-probe sets Ag2933, Ag3370 and Ag3837, described in Tables BDA, BDB and BDC.

Table BDA. Probe Name Ag2933

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-caccaccactaagccagaac-3'	20	4011	1237
Probe	TET-5'-ttctcagtcacaagaacatctcaaataatgt-3'-TAMRA	29	4031	1238
Reverse	5'-ggattcccatgtaattcaag-3'	21	4083	1239

Table BDB. Probe Name Ag3370

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agctggattcttccaaacaga-3'	21	1327	1240
Probe	TET-5'-tcacatgtatacatgctgccaatgg-3'- TAMRA	26	1375	1241
Reverse	5'-acctttgggatggaaagagtt-3'	21	1401	1242

Table BDC. Probe Name Ag3837

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acgagcttgaggatgtggat-3'	20	3725	1243
Probe	TET-5'-ttttgtcctctgtgacagtctccaca-3'- TAMRA	26	3758	1244
Reverse	5'-gcttcttctctgggtgaaatgg-3'	20	3784	1245

5 **CNS_neurodegeneration_v1.0 Summary:** Ag2933/Ag3370 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3370/Ag3837 The amp plots suggest that there were experimental difficulties with these runs (data not shown).

10 **Panel 1.3D Summary:** Ag2933 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 2D Summary: Ag2933 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag2933/Ag3370 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

15 **BE. CG56781-01: NEUROPSIN PRECURSOR**

Expression of gene CG56781-01 was assessed using the primer-probe sets Ag3019 and Ag4966, described in Tables BEA and BEB.

Table BEA. Probe Name Ag3019

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aggatctgagcctgtgttcag-3'	21	714	1246
Probe	TET-5'-cggagaccgcgtgtctacactaacgt-3'- TAMRA	26	739	1247
Reverse	5'-ttcaatccactccaggtagtca-3'	22	768	1248

20 **Table BEB. Probe Name Ag4966**

Primers	Sequences	Length	Start	SEQ ID NO:
---------	-----------	--------	-------	------------

			Position	
Forward	5'-ccacctcttctctcagag-3'	18	7	1249
Probe	TET-5'-caccctgtgcaatccagccgtg-3'- TAMRA	22	44	1250
Reverse	5'-acacctgcccacgctc-3'	16	89	1251

CNS_neurodegeneration_v1.0 Summary: Ag3019 The amp plot suggests that there were experimental difficulties with this run in one sample (data not shown). Given the lack of expression of this gene on the other panels the expression detected in the occipital cortex is likely artifactual.

Panel 1.3D Summary: Ag3019 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag3019 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3019 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 5 Islet Summary: Ag3019 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

NOV63

Expression of gene NOV63 was assessed using the primer-probe sets Ag2261 and Ag3035, described in Tables BFA and BFB. Results of the RTQ-PCR runs are shown in Tables BFC, BFD, BFE and BFF.

Table BFA. Probe Name Ag2261

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggatgactcgcttagcttct-3'	20	882	1252
Probe	TET-5'-gccgtaggtgccaccgtgagaag-3'- TAMRA	23	935	1253
Reverse	5'-agcagatgctctcgagtt-3'	19	958	1254

Table BFB. Probe Name Ag3035

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acagcagcaagttcgtcaag-3'	20	527	1255
Probe	TET-5'-agacggtcaagcaaggatctgcgag-3'- TAMRA	25	559	1256
Reverse	5'-cacgaggttggttggaagt-3'	20	593	1257

Table BFC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2261, Run 150631675	Rel. Exp.(%) Ag2261, Run 152887692	Rel. Exp.(%) Ag3035, Run 167597764	Tissue Name	Rel. Exp.(%) Ag2261, Run 150631675	Rel. Exp.(%) Ag2261, Run 152887692	Rel. Exp.(%) Ag3035, Run 167597764
Liver adenocarcinoma	22.4	19.6	71.2	Kidney (fetal)	2.1	0.0	2.7
Pancreas	3.9	2.5	2.8	Renal ca. 786-0	0.0	0.0	0.0
Pancreatic ca. CAPAN 2	5.3	3.5	9.5	Renal ca. A498	10.2	5.3	9.2
Adrenal gland	2.1	0.6	2.0	Renal ca. RXF 393	0.0	0.0	0.0
Thyroid	7.0	9.8	3.9	Renal ca. ACHN	0.0	2.2	0.0
Salivary gland	1.9	2.1	4.2	Renal ca. UO-31	0.0	0.0	0.0
Pituitary gland	1.0	2.2	6.7	Renal ca. TK-10	0.0	0.0	0.0
Brain (fetal)	6.8	4.9	10.8	Liver	0.0	0.0	0.0
Brain (whole)	4.8	3.0	1.4	Liver (fetal)	7.6	0.0	0.0
Brain (amygdala)	4.6	5.3	1.5	Liver ca. (hepatoblast) HepG2	0.0	0.0	0.0
Brain (cerebellum)	1.6	1.6	2.0	Lung	14.3	15.8	9.2
Brain (hippocampus)	7.5	11.3	0.6	Lung (fetal)	15.1	15.4	7.4
Brain (substantia nigra)	1.2	2.6	1.3	Lung ca. (small cell) LX-1	1.6	0.0	0.0
Brain (thalamus)	2.5	1.7	2.6	Lung ca. (small cell) NCI-H69	29.5	19.1	31.2
Cerebral Cortex	0.0	0.0	5.0	Lung ca. (s.cell var.) SHP-77	11.0	5.1	37.4
Spinal cord	1.7	2.1	2.7	Lung ca. (large cell) NCI-	0.0	0.0	0.0

				H460			
glio/astro U87-MG	0.0	0.0	0.0	Lung ca. (non-sm. cell) A549	0.0	1.2	1.6
glio/astro U-118-MG	55.1	50.3	42.9	Lung ca. (non- s.cell) NCI-H23	0.0	1.3	0.8
astrocytom a SW1783	0.0	7.5	0.0	Lung ca. (non- s.cell) HOP-62	0.0	1.7	0.0
neuro*; met SK-N- AS	0.0	0.0	0.7	Lung ca. (non-s.cl) NCI-H522	8.0	8.3	7.3
astrocytom a SF-539	1.9	4.7	9.9	Lung ca. (squam.) SW 900	4.0	0.0	1.8
astrocytom a SNB-75	2.0	4.9	6.9	Lung ca. (squam.) NCI-H596	15.8	10.2	58.2
glioma SNB-19	6.7	2.4	3.7	Mammary gland	7.2	4.1	4.4
glioma U251	2.1	4.5	6.8	Breast ca.* (pl.ef) MCF-7	1.7	3.4	7.3
glioma SF- 295	10.0	0.6	4.6	Breast ca.* (pl.ef) MDA- MB-231	23.2	19.6	19.2
Heart (fetal)	11.1	9.9	38.2	Breast ca.* (pl.ef) T47D	4.3	5.8	21.8
Heart	4.9	6.0	15.2	Breast ca. BT-549	0.0	4.2	2.2
Skeletal muscle (fetal)	100.0	100.0	85.3	Breast ca. MDA-N	0.0	0.0	0.0
Skeletal muscle	5.5	8.4	39.8	Ovary	3.6	3.1	8.1
Bone marrow	0.0	0.0	0.7	Ovarian ca. OVCAR-3	1.1	1.0	5.6
Thymus	10.0	3.9	6.4	Ovarian ca. OVCAR-4	0.0	0.0	0.7
Spleen	3.8	4.2	1.6	Ovarian	0.0	0.0	11.5

				ca. OVCAR-5			
Lymph node	5.0	1.1	1.4	Ovarian ca. OVCAR-8	1.3	4.3	4.1
Colorectal	3.4	5.4	6.8	Ovarian ca. IGROV-1	0.0	0.0	8.1
Stomach	6.0	15.4	3.1	Ovarian ca.* (ascites) SK-OV-3	7.5	16.0	100.0
Small intestine	15.9	18.7	2.3	Uterus	17.8	15.1	9.9
Colon ca. SW480	24.3	15.3	11.6	Placenta	4.6	8.2	2.1
Colon ca.* SW620(S W480 met)	0.0	0.0	2.1	Prostate	3.6	5.3	0.6
Colon ca. HT29	0.0	0.0	0.0	Prostate ca.* (bone met)PC-3	1.7	1.5	6.1
Colon ca. HCT-116	3.8	0.6	3.3	Testis	21.9	14.6	1.6
Colon ca. CaCo-2	0.0	0.8	0.3	Melanoma Hs688(A). T	3.1	4.7	1.4
Colon ca. tissue(OD O3866)	2.3	0.0	1.6	Melanoma * (met) Hs688(B). T	0.4	1.3	0.0
Colon ca. HCC-2998	0.0	0.0	1.6	Melanoma UACC-62	0.0	0.0	0.0
Gastric ca.* (liver met) NCI- N87	16.7	14.9	15.3	Melanoma M14	0.0	0.0	0.0
Bladder	1.6	3.2	3.0	Melanoma LOX IMVI	0.0	0.0	0.0
Trachea	24.3	33.7	5.7	Melanoma * (met) SK-MEL- 5	0.0	2.0	0.7
Kidney	0.0	0.0	0.0	Adipose	6.7	7.2	21.2

Table BFD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2261, Run 150811744	Rel. Exp.(%) Ag2261, Run 152887693	Tissue Name	Rel. Exp.(%) Ag2261, Run 150811744	Rel. Exp.(%) Ag2261, Run 152887693
Normal Colon	19.1	19.8	Kidney Margin 8120608	2.4	0.0
CC Well to Mod Diff (ODO3866)	0.0	5.8	Kidney Cancer 8120613	14.6	7.3
CC Margin (ODO3866)	19.5	12.5	Kidney Margin 8120614	4.8	1.5
CC Gr.2 rectosigmoid (ODO3868)	3.8	1.4	Kidney Cancer 9010320	0.0	0.0
CC Margin (ODO3868)	2.6	5.1	Kidney Margin 9010321	0.0	0.0
CC Mod Diff (ODO3920)	6.0	2.9	Normal Uterus	9.7	2.8
CC Margin (ODO3920)	23.8	6.4	Uterus Cancer 064011	85.9	41.5
CC Gr.2 ascend colon (ODO3921)	9.3	2.2	Normal Thyroid	15.2	7.3
CC Margin (ODO3921)	16.8	11.7	Thyroid Cancer 064010	0.0	3.0
CC from Partial Hepatectomy (ODO4309) Mets	2.4	0.0	Thyroid Cancer A302152	1.9	1.2
Liver Margin (ODO4309)	2.6	0.0	Thyroid Margin A302153	2.6	2.8
Colon mets to lung (OD04451- 01)	7.9	4.5	Normal Breast	16.2	2.7
Lung Margin (OD04451-02)	11.3	12.9	Breast Cancer (OD04566)	78.5	29.7
Normal Prostate 6546-1	6.3	2.6	Breast Cancer (OD04590-01)	37.6	23.8
Prostate Cancer (OD04410)	17.8	7.3	Breast Cancer Mets (OD04590-03)	100.0	24.5
Prostate Margin (OD04410)	10.7	7.4	Breast Cancer Metastasis	94.0	45.4

Kidney Margin (OD04338)	4.0	4.6	Normal Ovary	13.6	12.4
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	3.3	Ovarian Cancer 064008	37.9	2.1
Kidney Margin (OD04339)	18.7	0.0	Ovarian Cancer (OD04768-07)	18.4	3.7
Kidney Ca, Clear cell type (OD04340)	8.8	11.7	Ovary Margin (OD04768-08)	28.3	12.2
Kidney Margin (OD04340)	0.0	2.0	Normal Stomach	48.3	17.3
Kidney Ca, Nuclear grade 3 (OD04348)	3.5	4.0	Gastric Cancer 9060358	0.0	0.0
Kidney Margin (OD04348)	2.0	1.7	Stomach Margin 9060359	9.9	3.0
Kidney Cancer (OD04622-01)	9.3	0.0	Gastric Cancer 9060395	20.7	10.4
Kidney Margin (OD04622-03)	0.0	6.3	Stomach Margin 9060394	10.0	12.2
Kidney Cancer (OD04450-01)	0.0	0.0	Gastric Cancer 9060397	8.7	1.5
Kidney Margin (OD04450-03)	0.0	0.0	Stomach Margin 9060396	7.5	6.2
Kidney Cancer 8120607	0.0	0.7	Gastric Cancer 064005	10.7	4.8

Table BFE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3035, Run 190944495	Tissue Name	Rel. Exp.(%) Ag3035, Run 190944495
Secondary Th1 act	0.0	HUVEC IL-1beta	1.7
Secondary Th2 act	0.0	HUVEC IFN gamma	0.8
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.2
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.6
Secondary Th2 rest	0.0	HUVEC IL-11	1.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	2.7
Primary Th1 act	0.0	Lung Microvascular EC	0.6

		TNFalpha + IL-1beta	
Primary Th2 act	0.0	Microvascular Dermal EC none	3.8
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	1.2
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	3.7
Primary Th2 rest	0.0	Small airway epithelium none	1.9
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	4.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.2
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	2.4
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	1.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	2.1
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	22.2
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	18.8
LAK cells IL-2	0.0	Liver cirrhosis	0.7
LAK cells IL-2+IL-12	0.0	NCI-H292 none	0.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	1.5
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	2.0
LAK cells PMA/ionomycin	11.0	NCI-H292 IL-13	1.4
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	1.5
Two Way MLR 3 day	0.0	HPAEC none	3.1
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.5
Two Way MLR 7 day	0.0	Lung fibroblast none	6.2
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	2.1
PBMC PWM	0.5	Lung fibroblast IL-4	4.2
PBMC PHA-L	0.4	Lung fibroblast IL-9	8.3
Ramos (B cell) none	0.0	Lung fibroblast IL-13	4.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	8.1

B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	0.4
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	2.9
EOL-1 dbcAMP PMA/ionomycin	1.0	Dermal fibroblast IFN gamma	5.8
Dendritic cells none	0.0	Dermal fibroblast IL-4	17.2
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	4.8
Dendritic cells anti- CD40	0.0	Neutrophils TNFa+LPS	1.0
Monocytes rest	0.0	Neutrophils rest	2.2
Monocytes LPS	0.6	Colon	2.6
Macrophages rest	0.0	Lung	8.8
Macrophages LPS	0.0	Thymus	17.1
HUVEC none	2.4	Kidney	100.0
HUVEC starved	8.8		

Table BFF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2261, Run 152887762	Rel. Exp.(%) Ag3035, Run 165242424	Tissue Name	Rel. Exp.(%) Ag2261, Run 152887762	Rel. Exp.(%) Ag3035, Run 165242424
Secondary Th1 act	0.0	2.1	HUVEC IL-1beta	0.0	1.7
Secondary Th2 act	0.0	0.0	HUVEC IFN gamma	3.7	11.5
Secondary Tr1 act	0.0	4.2	HUVEC TNF alpha + IFN gamma	0.0	3.1
Secondary Th1 rest	0.0	0.0	HUVEC TNF alpha + IL4	4.3	5.1
Secondary Th2 rest	0.0	2.3	HUVEC IL-11	4.0	11.2
Secondary Tr1 rest	0.0	0.0	Lung Microvascular EC none	7.2	8.1
Primary Th1 act	0.0	0.0	Lung Microvascular EC TNFalpha + IL- 1beta	0.0	0.0
Primary Th2 act	0.0	0.0	Microvascular Dermal EC none	8.4	14.5
Primary Tr1 act	0.0	0.0	Microvascular Dermal EC	0.0	2.2

			TNFalpha + IL-1beta		
Primary Th1 rest	0.0	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0	16.3
Primary Th2 rest	0.0	0.0	Small airway epithelium none	5.9	18.8
Primary Tr1 rest	0.0	0.0	Small airway epithelium TNFalpha + IL-1beta	24.3	58.6
CD45RA CD4 lymphocyte act	0.0	0.0	Coronary artery SMC rest	0.0	2.0
CD45RO CD4 lymphocyte act	0.0	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0	0.0
CD8 lymphocyte act	0.0	0.0	Astrocytes rest	3.3	13.5
Secondary CD8 lymphocyte rest	0.0	0.7	Astrocytes TNFalpha + IL-1beta	0.0	8.6
Secondary CD8 lymphocyte act	1.6	0.0	KU-812 (Basophil) rest	0.0	0.0
CD4 lymphocyte none	0.0	0.0	KU-812 (Basophil) PMA/ionomycin	0.0	9.7
2ry Th1/Th2/Tr1 anti-CD95 CH11	0.0	1.4	CCD1106 (Keratinocytes) none	47.3	100.0
LAK cells rest	3.5	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	9.0	53.6
LAK cells IL-2	0.0	0.0	Liver cirrhosis	32.8	9.4
LAK cells IL-2+IL-12	0.0	0.0	Lupus kidney	0.0	1.6
LAK cells IL-2+IFN gamma	0.0	4.0	NCI-H292 none	3.8	3.4
LAK cells IL-2+IL-18	0.0	0.0	NCI-H292 IL-4	8.0	19.5
LAK cells PMA/ionomycin	26.1	50.7	NCI-H292 IL-9	0.0	4.2
NK Cells IL-2 rest	0.0	0.0	NCI-H292 IL-13	13.8	7.0
Two Way MLR 3 day	0.0	0.0	NCI-H292 IFN gamma	16.2	5.7
Two Way MLR 5	0.0	0.0	HPAEC none	6.7	30.1

day					
Two Way MLR 7 day	0.0	0.0	HPAEC TNF alpha + IL-1 beta	0.0	0.0
PBMC rest	0.0	0.0	Lung fibroblast none	7.6	42.0
PBMC PWM	0.0	0.0	Lung fibroblast TNF alpha + IL-1 beta	3.1	6.3
PBMC PHA-L	0.0	0.0	Lung fibroblast IL-4	4.3	34.2
Ramos (B cell) none	0.0	0.0	Lung fibroblast IL-9	12.7	27.5
Ramos (B cell) ionomycin	0.0	0.0	Lung fibroblast IL-13	6.8	19.9
B lymphocytes PWM	0.0	0.0	Lung fibroblast IFN gamma	30.4	51.1
B lymphocytes CD40L and IL-4	3.1	0.0	Dermal fibroblast CCD1070 rest	0.0	2.8
EOL-1 dbcAMP	0.0	0.0	Dermal fibroblast CCD1070 TNF alpha	5.2	19.6
EOL-1 dbcAMP PMA/ionomycin	3.5	2.7	Dermal fibroblast CCD1070 IL-1 beta	0.0	2.0
Dendritic cells none	0.0	0.0	Dermal fibroblast IFN gamma	28.5	32.1
Dendritic cells LPS	0.0	0.0	Dermal fibroblast IL-4	42.9	91.4
Dendritic cells anti-CD40	0.0	0.0	IBD Colitis 2	2.2	5.5
Monocytes rest	0.0	0.0	IBD Crohn's	3.1	9.6
Monocytes LPS	0.0	0.0	Colon	100.0	58.6
Macrophages rest	0.0	0.0	Lung	36.3	26.1
Macrophages LPS	0.0	0.0	Thymus	0.0	0.0
HUVEC none	0.0	17.7	Kidney	4.0	33.0
HUVEC starved	17.4	51.1			

Panel 1.3D Summary: Ag2261 The NOV63 gene is expressed at moderate levels in a number of metabolic tissues, with highest overall expression seen in fetal skeletal muscle (CTs=30.4-31.8). The higher levels of expression in fetal skeletal muscle when compared to adult skeletal muscle suggests that the protein product encoded by the 88091010_EXT gene may be useful in treating muscular dystrophy, Lesch-Nyhan syndrome, myasthenia gravis and other conditions that result in weak or dystrophic muscle. This gene is also expressed in

adipose, thyroid and heart. Since biologic cross-talk between adipose and thyroid is a component of some forms of obesity, this gene product may be a protein therapeutic for the treatment of metabolic disease, including obesity and Type 2 diabetes.

5 Ag3035 This probe/primer set recognizes a distinct portion of this gene and shows a distinctive expression pattern when compared to Ag2261. This observation may indicate that the probe/primer sets can distinguish splice variants of this gene. In contrast to the results obtained with Ag 2261, expression of this gene is highest in an ovarian cancer cell line (CT = 30.6). As is the case for Ag2261, expression of this gene using Ag3035 also shows relatively high levels in fetal skeletal muscle. However, in addition, Ag3035 shows increased levels of
10 this gene in adult skeletal muscle as well as in adult and fetal heart. Most other expression is similar using both probe/primer sets. Please see Ag2261 for additional information.

Panel 2D Summary: Ag2261 The expression of this gene was assessed in two independent runs on panel 2D. This is consistently expressed in samples of breast cancer, uterine cancer and lung cancer when compared to their respective normal adjacent tissue
15 controls. Thus, the expression of this gene could be used to distinguish breast cancer, lung cancer or uterine cancer from their normal tissues. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, antibodies or protein therapeutics might be of use in the treatment of breast, lung or uterine cancer.

Panel 4.1D Summary: Ag3035 This probe/primer set recognizes a distinct portion of
20 this gene and shows a distinctive expression pattern when compared to Ag2261 in Panel 4D. This observation may indicate that the probe/primer sets can distinguish splice variants of this gene. In contrast to the results obtained with Ag2261, expression of this gene is highest in kidney (CT = 30.6). Most other expression is similar using both probe/primer sets. The NOV63 gene, a WNT-14 homolog is also expressed at moderate to low levels in several
25 unstimulated or cytokine-activated keratinocyte and lung and dermal fibroblast preparations (CT range 29-34). Thus, the NOV63 gene may be useful as a protein therapeutic that reduces or eliminates the symptoms of chronic obstructive pulmonary disease, asthma, emphysema, or psoriasis. In addition, due to its known effects on development of vertebrate joints, the protein encoded by the NOV63 gene may also reduce or eliminate the symptoms of osetoarthritis (See
30 Hartmann and Tabin, 2001).

References:

Christine Hartmann and Clifford J. Tabin Wnt-14 Plays a Pivotal Role in Inducing Synovial Joint Formation in the Developing Appendicular Skeleton Cell, Vol 104, 341-351, February 2001

The long bones of the vertebrate appendicular skeleton arise from initially continuous condensations of mesenchymal cells that subsequently segment and cavitate to form discrete elements separated by synovial joints. Little is known, however, about the molecular mechanisms of joint formation. We present evidence that Wnt-14 plays a central role in initiating synovial joint formation in the chick limb. Wnt-14 is expressed in joint-forming regions prior to the segmentation of the cartilage elements, and local misexpression of Wnt-14 induces morphological and molecular changes characteristic of the first steps of joint formation. Induction of an ectopic joint-like region by Wnt-14 suppresses the formation of the immediately adjacent endogenous joint, potentially providing insight into the spacing of joints.

Panel 4D Summary: Ag2261 The NOV63 transcript is expressed at low levels in colon (CT=33.5). Low but significant levels of expression are also found in the lung, keratinocytes and dermal fibroblast. Thus, this transcript could be used as a marker for thymic, lung and skin tissues. The putative Wnt-14 molecule encoded by this transcript may play an important role in the normal homeostasis of these tissues. Therapeutics designed with the protein encoded by this transcript could be important for maintaining or restoring normal function to these organs during inflammation.

NOV64

Expression of gene NOV64 was assessed using the primer-probe set Ag3043, described in Table BGA. Results of the RTQ-PCR runs are shown in Tables BGB, BGC and BGD.

Table BGA. Probe Name Ag3043

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cctgtatgaggaagtcgatgag-3'	22	868	1258
Probe	TET-5'-aggtcattcacgtccctctctctg-3'-TAMRA	24	900	1259
Reverse	5'-gatacgagtcctgtcttcctttc-3'	22	932	1260

Table BGB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3043, Run 211012232	Tissue Name	Rel. Exp.(%) Ag3043, Run 211012232
AD 1 Hippo	23.0	Control (Path) 3 Temporal Ctx	10.2
AD 2 Hippo	32.8	Control (Path) 4 Temporal Ctx	28.3
AD 3 Hippo	12.3	AD 1 Occipital Ctx	20.2

AD 4 Hippo	11.2	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	88.9	AD 3 Occipital Ctx	13.1
AD 6 Hippo	73.2	AD 4 Occipital Ctx	26.2
Control 2 Hippo	31.6	AD 5 Occipital Ctx	47.6
Control 4 Hippo	15.0	AD 6 Occipital Ctx	27.5
Control (Path) 3 Hippo	13.1	Control 1 Occipital Ctx	8.7
AD 1 Temporal Ctx	32.3	Control 2 Occipital Ctx	57.0
AD 2 Temporal Ctx	44.4	Control 3 Occipital Ctx	25.5
AD 3 Temporal Ctx	15.4	Control 4 Occipital Ctx	11.9
AD 4 Temporal Ctx	29.1	Control (Path) 1 Occipital Ctx	68.8
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	14.1
AD 5 Sup Temporal Ctx	54.3	Control (Path) 3 Occipital Ctx	6.7
AD 6 Inf Temporal Ctx	72.7	Control (Path) 4 Occipital Ctx	17.3
AD 6 Sup Temporal Ctx	58.2	Control 1 Parietal Ctx	12.9
Control 1 Temporal Ctx	12.0	Control 2 Parietal Ctx	47.0
Control 2 Temporal Ctx	37.1	Control 3 Parietal Ctx	19.6
Control 3 Temporal Ctx	21.0	Control (Path) 1 Parietal Ctx	62.4
Control 3 Temporal Ctx	13.5	Control (Path) 2 Parietal Ctx	24.8
Control (Path) 1 Temporal Ctx	56.6	Control (Path) 3 Parietal Ctx	6.0
Control (Path) 2 Temporal Ctx	40.3	Control (Path) 4 Parietal Ctx	46.3

Table BGC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3043, Run 167963717	Tissue Name	Rel. Exp.(%) Ag3043, Run 167963717
Liver adenocarcinoma	79.0	Kidney (fetal)	21.6
Pancreas	4.3	Renal ca. 786-0	19.2
Pancreatic ca. CAPAN 2	16.3	Renal ca. A498	18.4

		5	
Lymph node	19.3	Ovarian ca. OVCAR-8	6.3
Colorectal	12.9	Ovarian ca. IGROV-1	6.5
Stomach	3.7	Ovarian ca.* (ascites) SK-OV-3	100.0
Small intestine	5.4	Uterus	5.8
Colon ca. SW480	14.2	Placenta	3.8
Colon ca.* SW620(SW480 met)	45.4	Prostate	2.9
Colon ca. HT29	12.7	Prostate ca.* (bone met)PC-3	40.9
Colon ca. HCT-116	25.9	Testis	2.8
Colon ca. CaCo-2	25.2	Melanoma Hs688(A).T	6.8
Colon ca. tissue(ODO3866)	15.4	Melanoma* (met) Hs688(B).T	7.7
Colon ca. HCC-2998	13.3	Melanoma UACC-62	17.3
Gastric ca.* (liver met) NCI-N87	10.4	Melanoma M14	11.0
Bladder	13.8	Melanoma LOX IMVI	43.5
Trachea	3.0	Melanoma* (met) SK-MEL-5	27.5
Kidney	9.5	Adipose	7.3

Table BGD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3043, Run 164315037	Tissue Name	Rel. Exp.(%) Ag3043, Run 164315037
Secondary Th1 act	38.4	HUVEC IL-1beta	20.6
Secondary Th2 act	38.7	HUVEC IFN gamma	22.5
Secondary Tr1 act	40.3	HUVEC TNF alpha + IFN gamma	28.7
Secondary Th1 rest	10.2	HUVEC TNF alpha + IL4	40.6
Secondary Th2 rest	18.2	HUVEC IL-11	18.4
Secondary Tr1 rest	16.3	Lung Microvascular EC none	23.3
Primary Th1 act	47.6	Lung Microvascular EC TNFalpha + IL-1beta	26.1
Primary Th2 act	34.6	Microvascular Dermal EC none	47.0
Primary Tr1 act	46.3	Microvascular Dermal EC	28.5

EOL-1 dbcAMP	24.7	Dermal fibroblast CCD1070 TNF alpha	84.7
EOL-1 dbcAMP PMA/ionomycin	40.9	Dermal fibroblast CCD1070 IL-1 beta	25.0
Dendritic cells none	34.2	Dermal fibroblast IFN gamma	29.3
Dendritic cells LPS	38.7	Dermal fibroblast IL-4	42.6
Dendritic cells anti- CD40	48.0	IBD Colitis 2	2.0
Monocytes rest	17.9	IBD Crohn's	1.8
Monocytes LPS	19.1	Colon	17.3
Macrophages rest	40.6	Lung	16.8
Macrophages LPS	18.4	Thymus	17.7
HUVEC none	37.4	Kidney	29.1
HUVEC starved	65.1		

CNS_neurodegeneration_v1.0 Summary: Ag3043 There is an association with a statistical confidence of 0.1 between increased expression of the NOV64 gene in the temporal cortex and Alzheimer's disease. This gene encodes a homolog of dipeptidyl peptidase, which belongs to a known class of markers of T cell activation in Multiple Sclerosis. This indicates that inhibitors of this gene product may have utility in treatment of this disease. A dipeptidyl peptidase is also dysregulated in Huntington's disease. Our finding of increased expression of this gene product in the temporal cortex of Alzheimer's disease patients indicates that there may be a wider utility of inhibitors of the protein encoded by this gene, including the treatment of neurodegenerative diseases such as Huntington's and Alzheimer's, as well as Multiple Sclerosis.

References:

Khoury SJ, Guttmann CR, Orav EJ, Kikinis R, Jolesz FA, Weiner HL. Changes in activated T cells in the blood correlate with disease activity in multiple sclerosis. Arch Neurol 2000 Aug;57(8):1183-9

OBJECTIVE: To determine whether changes in activation markers on peripheral blood T cells correlate with disease activity in patients with multiple sclerosis. **DESIGN:** In a prospective longitudinal study during 1 year, we analyzed the change in percentage of activated T lymphocytes in the peripheral blood of 40 patients with multiple sclerosis in relation to clinical findings and changes on brain magnetic resonance imaging (MRI) scans. The patients underwent repeated imaging of the brain (mean number of MRIs for each patient, 22) at the time blood samples were obtained as well as at monthly neurological examinations, and at the time of scoring on the Kurtzke Expanded Disability Status Scale (EDSS) and

ambulation index scale. RESULTS: A change in the percentage of cells expressing the activation markers interleukin 2 receptor (CD25), class II major histocompatibility complex (MHC) (I3) or surface dipeptidyl peptidase (CD26) correlated significantly with a change in lesion volume or a change in number of gadolinium-enhancing lesions as detected on MRI.

- 5 Changes in CD25(+) cells and in CD4(+) cells expressing class II MHC also correlated with changes in disability as measured by EDSS in patients with relapsing-remitting disease, and changes in CD4(+)CD25(+) cells correlated with the occurrence of attacks in patients with relapsing-remitting disease. These correlations are dependent on measurement of changes between time points sampled at 1- or 2-week intervals. CONCLUSION: There is a linkage
10 between peripheral T-lymphocyte activation as measured by cell surface markers and disease activity in patients with multiple sclerosis.

- Mantle D, Falkous G, Ishiura S, Perry RH, Perry EK. Comparison of cathepsin protease activities in brain tissue from normal cases and cases with Alzheimer's disease, Lewy body dementia, Parkinson's disease and Huntington's disease. *J Neurol Sci* 1995 Jul;131(1):65-
15 70

- Recent evidence, based upon immunocytochemical and histochemical analysis of brain cortical tissue from Alzheimer's disease patients, has suggested that altered activity and/or distribution of the lysosomal proteases cathepsins B and D may be implicated in the abnormal protein processing pathway resulting in formation of the neurotoxic amyloid A4 peptide,
20 characteristic of this neurodegenerative disorder. We have therefore compared, via biochemical assay techniques using conventional or specially synthesised (corresponding to protein cleavage points of relevant to A4 peptide formation) fluorogenic substrates, the levels of activity of the lysosomal proteases cathepsins B, D, H and L, and dipeptidyl aminopeptidases I and II in frontal cortex (grey/white matter) from control and Alzheimer's
25 disease patients. For comparative purposes, activity levels of the above enzymes were also determined in frontal cortex tissue from cases with Lewy body dementia and Parkinson's disease, and in caudate tissue from control and Huntington's disease cases. There was no significant difference in activity for any protease types in tissue from control cases and cases with Alzheimer's disease, Lewy body dementia or Parkinson's disease, with the exception of
30 reduced dipeptidyl aminopeptidase II activity in Lewy body dementia and Parkinson's cases. We have therefore been unable to confirm a potential role for lysosomal cathepsins in the characteristic neurodegeneration associated with Alzheimer's disease; however the finding of significant increases in activity of dipeptidyl aminopeptidase II, cathepsin H and cathepsin D specifically in cases with Huntington's disease is of particular note. We therefore suggest the

potential role of the latter enzymes in the pathogenesis of Huntington's disease requires further investigation

Panel 1.3D Summary: Ag3043 Highest expression of the NOV64 gene is seen in an ovarian cancer cell line (CT=26.2). This gene is expressed at moderate levels in all the cancer cell lines in this panel. Thus, this is a potential target for small molecule inhibitor drugs in cancer.

This gene also has moderate levels of expression in pancreas, adrenal, thyroid, pituitary, heart, skeletal muscle, liver and adipose. Therefore, this gene product may be a small molecule target for the treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene is expressed in the central nervous system. Please see CNS_neurodegeneration_v1.0 for discussion of utility of this gene in the central nervous system.

Panel 4D Summary: Ag3043 The NOV64 gene is expressed in a number of cells and tissues of immunological importance, especially in activated B cells, T cells, dendritic cells, and activated lung and skin fibroblasts. Therefore, small molecule antagonists that block the function of the NOV64 gene product may reduce or eliminate the symptoms of a wide range of autoimmune and inflammatory diseases, including Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

NOV65a and NOV65b

Expression of gene NOV65a and variant NOV65b was assessed using the primer-probe sets Ag3020 and Ag2968, described in Tables BHA and BHB. Results of the RTQ-PCR runs are shown in Tables BHC, BHD, BHE, BHF and BHG.

Table BHA. Probe Name Ag3020

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaatcaccacattctgaat-3'	21	193	1261
Probe	TET-5'-cgtttacactggccccgaattctaca-3'-TAMRA	26	231	1262
Reverse	5'-cctctacaccaggtactggat-3'	22	268	1263

Table BHB. Probe Name Ag2968

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaatcaccacattctgaat-3'	21	193	1264
Probe	TET-5'-cgtttacactggccccgaattctaca-3'-	26	231	1265

	TAMRA			
Reverse	5'-cctctacacccaggtactggat-3'	22	268	1266

Table BHC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3020, Run 221998694	Tissue Name	Rel. Exp.(%) Ag3020, Run 221998694
Adipose	0.4	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.2	Colon ca. CaCo-2	0.0
Placenta	0.1	Colon cancer tissue	31.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	100.0
Ovarian ca. SK-OV- 3	0.3	Colon ca. SW-48	5.7
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	0.0
Ovarian ca. IGROV- 1	18.9	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	12.7
Breast ca. MCF-7	0.0	Heart Pool	4.9
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	23.3
Breast ca. T47D	0.0	Skeletal Muscle Pool	20.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0

Trachea	0.3	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.1
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	0.7
Lung ca. LX-1	12.7	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	15.9
Lung ca. SHP-77	0.0	CNS cancer (glio) SF- 295	0.5
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.1
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	0.1
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.1
Liver	0.0	Brain (Thalamus) Pool	0.1
Fetal Liver	0.1	Brain (whole)	0.1
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.1	Adrenal Gland	0.0
Fetal Kidney	0.1	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.4
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table BHD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2968, Run 166220058	Rel. Exp.(%) Ag3020, Run 167819114	Tissue Name	Rel. Exp.(%) Ag2968, Run 166220058	Rel. Exp.(%) Ag3020, Run 167819114
Liver adenocarcinoma	0.0	0.1	Kidney (fetal)	0.0	0.3
Pancreas	0.0	0.0	Renal ca. 786- 0	0.0	0.0
Pancreatic ca. CAPAN 2	0.1	0.2	Renal ca. A498	0.0	0.0
Adrenal gland	0.0	0.0	Renal ca. RXF 393	0.0	0.0

Thyroid	0.3	1.6	Renal ca. ACHN	0.0	0.0
Salivary gland	0.8	0.6	Renal ca. UO-31	0.0	0.0
Pituitary gland	0.0	0.0	Renal ca. TK-10	0.0	0.0
Brain (fetal)	0.0	0.1	Liver	0.0	0.0
Brain (whole)	0.4	1.0	Liver (fetal)	0.5	0.0
Brain (amygdala)	0.2	1.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	0.0	0.0	Lung	0.0	0.5
Brain (hippocampus)	0.0	0.0	Lung (fetal)	0.0	0.0
Brain (substantia nigra)	0.2	0.1	Lung ca. (small cell) LX-1	10.7	16.2
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.5
Cerebral Cortex	0.4	0.1	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	0.0	0.0	Lung ca. (large cell) NCI-H460	0.0	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non-sm. cell) A549	0.0	0.2
glio/astro U-118-MG	0.0	0.1	Lung ca. (non-s.cell) NCI-H23	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non-s.cell) HOP-62	0.1	0.1
neuro*; met SK-N-AS	0.0	0.0	Lung ca. (non-s.cl) NCI-H522	0.0	0.0
astrocytoma SF-539	12.1	8.4	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SNB-75	0.0	0.3	Lung ca. (squam.) NCI-H596	0.1	0.4
glioma SNB-19	0.0	0.0	Mammary gland	0.2	0.2
glioma U251	0.4	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SF-295	0.3	0.3	Breast ca.* (pl.ef) MDA-	0.0	0.0

			MB-231		
Heart (fetal)	7.4	26.8	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	29.9	35.4	Breast ca. BT- 549	0.0	0.0
Skeletal muscle (fetal)	10.8	33.9	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	100.0	100.0	Ovary	0.0	0.1
Bone marrow	0.1	0.6	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	0.1	0.1	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.0	0.0	Ovarian ca. OVCAR-5	0.0	0.6
Lymph node	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.2	0.2	Ovarian ca. IGROV-1	26.2	26.2
Stomach	0.0	0.0	Ovarian ca.* (ascites) SK- OV-3	0.2	1.0
Small intestine	0.0	0.0	Uterus	0.0	0.0
Colon ca. SW480	0.0	0.2	Placenta	1.0	0.0
Colon ca.* SW620(SW480 met)	1.6	6.1	Prostate	0.2	0.1
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC- 3	0.0	0.0
Colon ca. HCT- 116	0.0	0.0	Testis	0.2	0.2
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	21.9	30.6	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC- 2998	0.0	0.5	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma M14	0.0	0.0
Bladder	0.1	0.0	Melanoma LOX IMVI	0.0	0.0
Trachea	0.8	0.6	Melanoma* (met) SK- MEL-5	0.0	0.0
Kidney	0.0	0.0	Adipose	1.1	1.7

Table BHE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2968, Run 170188142	Tissue Name	Rel. Exp.(%) Ag2968, Run 170188142
Daoy- Medulloblastoma	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	2.2	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	0.0	JM1- pre-B-cell lymphoma	0.0
Cerebellum	0.0	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell	0.0	Hs766T- Pancreatic	0.0

lung cancer		carcinoma (LN metastasis)	
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.0
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0
LX-1- Small cell lung cancer	4.4	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	100.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	4.3	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	10.1	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	1.2	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.7
NCI-SNU-1- Gastric carcinoma	0.1	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0

RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table BHF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3020, Run 164528102	Tissue Name	Rel. Exp.(%) Ag3020, Run 164528102
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	65.1
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	21.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	10.7
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	13.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0

LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	8.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	10.6	HPAEC TNF alpha + IL-1 beta	11.2
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	9.9	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	23.8	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	44.4
Macrophages rest	0.0	Lung	26.6
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Table BHG. Panel 5D

Tissue Name	Rel. Exp.(%) Ag3020, Run 172171108	Tissue Name	Rel. Exp.(%) Ag3020, Run 172171108
-------------	--	-------------	--

97457_Patient-02go_adipose	0.2	94709_Donor 2 AM - A_adipose	0.0
97476_Patient-07sk_skeletal muscle	8.4	94710_Donor 2 AM - B_adipose	0.0
97477_Patient-07ut_uterus	0.0	94711_Donor 2 AM - C_adipose	0.0
97478_Patient-07pl_placenta	1.3	94712_Donor 2 AD - A_adipose	3.0
97481_Patient-08sk_skeletal muscle	12.6	94713_Donor 2 AD - B_adipose	0.0
97482_Patient-08ut_uterus	0.0	94714_Donor 2 AD - C_adipose	0.0
97483_Patient-08pl_placenta	0.0	94742_Donor 3 U - A_Mesenchymal Stem Cells	0.0
97486_Patient-09sk_skeletal muscle	12.9	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0
97487_Patient-09ut_uterus	0.7	94730_Donor 3 AM - A_adipose	0.0
97488_Patient-09pl_placenta	0.2	94731_Donor 3 AM - B_adipose	0.0
97492_Patient-10ut_uterus	0.2	94732_Donor 3 AM - C_adipose	0.0
97493_Patient-10pl_placenta	1.1	94733_Donor 3 AD - A_adipose	0.0
97495_Patient-11go_adipose	0.0	94734_Donor 3 AD - B_adipose	0.0
97496_Patient-11sk_skeletal muscle	53.6	94735_Donor 3 AD - C_adipose	0.0
97497_Patient-11ut_uterus	0.3	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	2.3	73556_Heart_Cardiac stromal cells (primary)	0.0
97500_Patient-12go_adipose	0.4	81735_Small Intestine	0.0
97501_Patient-12sk_skeletal muscle	100.0	72409_Kidney_Proximal Convoluted Tubule	0.0
97502_Patient-12ut_uterus	0.0	82685_Small intestine_Duodenum	0.0
97503_Patient-12pl_placenta	1.3	90650_Adrenal_Adrenocortical adenoma	0.2
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	72410_Kidney_HRCE	0.3
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	72411_Kidney_HRE	0.0

94723_Donor 2 U - C_Mesenchymal Stem Cells	0.0	73139_Uterus_Uterine smooth muscle cells	0.0
--	-----	---	-----

General_screening_panel_v1.4 Summary: Ag3020 The NOV65a gene is expressed in brain, colon, lung and ovarian cancer cell lines with highest expression in a colon cancer cell line Colo-205 (CT=24.37). This suggests that this gene can be used as a diagnostic marker for these types of cancer . Furthermore, inhibition of the protein using small molecule drugs could potentially be useful for the treatment of brain, colon, lung and ovarian cancer.

In addition, this gene has low expression in adipose and high expression in adult and fetal heart and skeletal muscle. Thus, this protein phosphatase may be a small molecule target for the treatment of obesity, Type 2 diabetes and cardiac and skeletal muscle disease.

Panel 1.3D Summary: Ag2968/Ag3020 Results from two experiments using identical probe/primer sets are in excellent agreement. Expression of the NOV65a gene is highest in adult skeletal muscle (CTs = 26-28). Significant but somewhat lower expression is also seen in fetal skeletal muscle and adult/fetal heart. Thus, expression of this gene may be used to distinguish these samples from the other samples on this panel.

This gene is also expressed in brain, colon, lung and ovarian cancer cell lines, consistent with General_screening_panel_v1.4. This suggests that this gene can be used as a diagnostic marker for these types of cancer and inhibition of the protein using small molecule drugs can be used for the treatment of brain, colon, lung and ovarian cancer.

Panel 3D Summary: Ag2968 Expression of the NOV65a gene is highest in colon cancer cell line Colo-205 (CT = 25.6). In addition, significant expression of this gene is seen in two other colon cancer cell lines. Thus, expression of this gene may be used to distinguish these colon cancer cell lines from the other samples on this panel. Moreover, therapeutic modulation of the activity of this gene or its protein product, using small molecules, antibodies or protein therapeutics, may be of benefit in the treatment of colon cancer.

Panel 4D Summary: Ag3020 Expression of the NOV65a gene is highest in a liver cirrhosis sample (CT = 33.3). Furthermore, expression of this gene is not detected in normal liver in Panels 1.3D or 1.4, suggesting that its expression is unique to liver cirrhosis. This gene encodes a putative protein phosphatase; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this protein could also be used for the diagnosis of liver cirrhosis. Low levels of expression are also seen in colon and resting astrocytes.

Ag2968 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 5D Summary: Ag3020 Expression of the NOV65a gene is primarily restricted to samples from skeletal muscle. This specific expression is in agreement with the results in Panels 1.3D and 1.4. Thus, expression of this gene could be used to differentiate between skeletal muscle and other samples on this panel, and as a marker of skeletal muscle. Results from one experiment with the probe and primer set Ag2968 are not included. The amp plot indicates that there were experimental difficulties with this run.

NOV66

Expression of gene NOV66 was assessed using the primer-probe set Ag2913, described in Table BIA. Results of the RTQ-PCR runs are shown in Tables BIB, BIC and BID.

Table BIA. Probe Name Ag2913

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tttgtggcttgatggcttt-3'	19	956	1267
Probe	TET-5'-ttcctttccgcatttcctatgtgaat-3'-TAMRA	26	977	1268
Reverse	5'-ttccagttaaaggcataacgaa-3'	22	1012	1269

15

Table BIB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2913, Run 157366466	Tissue Name	Rel. Exp.(%) Ag2913, Run 157366466
Liver adenocarcinoma	1.2	Kidney (fetal)	4.3
Pancreas	0.4	Renal ca. 786-0	0.6
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.6
Adrenal gland	2.3	Renal ca. RXF 393	0.2
Thyroid	0.7	Renal ca. ACHN	0.0
Salivary gland	8.3	Renal ca. UO-31	0.0
Pituitary gland	0.3	Renal ca. TK-10	0.0
Brain (fetal)	26.4	Liver	2.0
Brain (whole)	6.4	Liver (fetal)	3.5
Brain (amygdala)	18.4	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	8.6	Lung	9.2
Brain (hippocampus)	100.0	Lung (fetal)	4.7
Brain (substantia nigra)	2.6	Lung ca. (small cell)	0.6

		LX-1	
Brain (thalamus)	9.5	Lung ca. (small cell) NCI-H69	2.1
Cerebral Cortex	31.2	Lung ca. (s.cell var.) SHP-77	3.8
Spinal cord	1.0	Lung ca. (large cell)NCI-H460	1.5
glio/astro U87-MG	1.2	Lung ca. (non-sm. cell) A549	2.2
glio/astro U-118-MG	18.9	Lung ca. (non-s.cell) NCI-H23	7.2
astrocytoma SW1783	2.8	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	8.2	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	2.5	Lung ca. (squam.) SW 900	0.9
astrocytoma SNB-75	0.7	Lung ca. (squam.) NCI-H596	0.8
glioma SNB-19	1.6	Mammary gland	5.7
glioma U251	0.4	Breast ca.* (pl.ef) MCF-7	2.2
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	3.1
Heart (fetal)	0.1	Breast ca.* (pl.ef) T47D	1.4
Heart	0.4	Breast ca. BT-549	2.6
Skeletal muscle (fetal)	0.4	Breast ca. MDA-N	1.8
Skeletal muscle	0.5	Ovary	0.0
Bone marrow	8.5	Ovarian ca. OVCAR- 3	0.9
Thymus	2.6	Ovarian ca. OVCAR- 4	0.0
Spleen	7.7	Ovarian ca. OVCAR- 5	0.0
Lymph node	9.1	Ovarian ca. OVCAR- 8	1.0
Colorectal	8.4	Ovarian ca. IGROV- 1	0.0
Stomach	0.9	Ovarian ca.* (ascites) SK-OV-3	0.9
Small intestine	11.2	Uterus	1.2
Colon ca. SW480	0.0	Placenta	1.7
Colon ca.* SW620(SW480 met)	0.0	Prostate	1.9

Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.1	Testis	2.8
Colon ca. CaCo-2	7.7	Melanoma Hs688(A).T	0.2
Colon ca. tissue(ODO3866)	1.1	Melanoma* (met) Hs688(B).T	0.8
Colon ca. HCC-2998	7.9	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	9.2	Melanoma M14	0.7
Bladder	6.4	Melanoma LOX IMVI	1.5
Trachea	1.4	Melanoma* (met) SK-MEL-5	1.3
Kidney	1.4	Adipose	2.3

Table BIC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2913, Run 157366467	Tissue Name	Rel. Exp.(%) Ag2913, Run 157366467
Normal Colon	32.5	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	11.2	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	8.7	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	13.5	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	9.5	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	35.4	Normal Uterus	0.0
CC Margin (ODO3920)	57.4	Uterus Cancer 064011	1.6
CC Gr.2 ascend colon (ODO3921)	30.8	Normal Thyroid	11.8
CC Margin (ODO3921)	6.4	Thyroid Cancer 064010	0.3
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	59.0
Liver Margin (ODO4309)	5.3	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451-02)	0.4	Breast Cancer (OD04566)	0.0

Normal Prostate 6546-1	1.6	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	21.8	Breast Cancer Mets (OD04590-03)	0.4
Prostate Margin (OD04410)	19.6	Breast Cancer Metastasis (OD04655-05)	1.3
Prostate Cancer (OD04720-01)	16.2	Breast Cancer 064006	2.4
Prostate Margin (OD04720-02)	18.9	Breast Cancer 1024	0.0
Normal Lung 061010	73.7	Breast Cancer 9100266	10.7
Lung Met to Muscle (ODO4286)	4.0	Breast Margin 9100265	2.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	2.4
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	1.9
Lung Margin (OD03126)	2.1	Normal Liver	2.6
Lung Cancer (OD04404)	0.4	Liver Cancer 064003	4.5
Lung Margin (OD04404)	0.0	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD04565)	2.5	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	5.2	Liver Tissue 6004-N	1.5
Lung Margin (OD04237-02)	1.9	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.4	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	2.7	Normal Bladder	18.6
Melanoma Mets to Lung (OD04321)	0.7	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	12.6
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	6.3
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	Bladder Normal Adjacent (OD04718-03)	7.3
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.5	Ovarian Cancer (OD04768-07)	0.0

CD45RO CD4 lymphocyte act	16.7	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	15.5	Astrocytes rest	0.6
Secondary CD8 lymphocyte rest	17.9	Astrocytes TNFalpha + IL-1beta	0.5
Secondary CD8 lymphocyte act	16.6	KU-812 (Basophil) rest	4.6
CD4 lymphocyte none	16.2	KU-812 (Basophil) PMA/ionomycin	18.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	29.9	CCD1106 (Keratinocytes) none	7.2
LAK cells rest	28.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.2
LAK cells IL-2	18.9	Liver cirrhosis	8.1
LAK cells IL-2+IL-12	6.0	Lupus kidney	3.4
LAK cells IL-2+IFN gamma	6.8	NCI-H292 none	4.3
LAK cells IL-2+ IL-18	3.5	NCI-H292 IL-4	2.1
LAK cells PMA/ionomycin	8.8	NCI-H292 IL-9	1.8
NK Cells IL-2 rest	4.6	NCI-H292 IL-13	0.2
Two Way MLR 3 day	12.4	NCI-H292 IFN gamma	0.5
Two Way MLR 5 day	6.9	HPAEC none	2.6
Two Way MLR 7 day	5.1	HPAEC TNF alpha + IL-1 beta	6.3
PBMC rest	4.9	Lung fibroblast none	9.9
PBMC PWM	37.4	Lung fibroblast TNF alpha + IL-1 beta	2.5
PBMC PHA-L	12.9	Lung fibroblast IL-4	12.7
Ramos (B cell) none	11.3	Lung fibroblast IL-9	8.7
Ramos (B cell) ionomycin	13.7	Lung fibroblast IL-13	12.1
B lymphocytes PWM	1.1	Lung fibroblast IFN gamma	9.3
B lymphocytes CD40L and IL-4	0.5	Dermal fibroblast CCD1070 rest	9.9
EOL-1 dbcAMP	0.4	Dermal fibroblast CCD1070 TNF alpha	29.3
EOL-1 dbcAMP PMA/ionomycin	0.9	Dermal fibroblast CCD1070 IL-1 beta	0.7
Dendritic cells none	0.7	Dermal fibroblast IFN gamma	1.4
Dendritic cells LPS	1.5	Dermal fibroblast IL-4	5.1
Dendritic cells anti-CD40	4.7	IBD Colitis 2	16.6

Monocytes rest	5.6	IBD Crohn's	7.9
Monocytes LPS	0.3	Colon	24.7
Macrophages rest	3.6	Lung	13.4
Macrophages LPS	0.4	Thymus	72.7
HUVEC none	1.2	Kidney	65.1
HUVEC starved	1.2		

CNS_neurodegeneration_v1.0 Summary: Ag2913 No significant expression detected. Potential failed chemistry reaction or bad probe/primer set (data not shown).

Panel 1.3D Summary: Ag2913 The NOV66 gene represents a novel G-protein coupled receptor (GPCR) with expression in the brain. The GPCR family of receptors contains a large number of neurotransmitter receptors, including the dopamine, serotonin, α and β -adrenergic, acetylcholine muscarinic, histamine, peptide, and metabotropic glutamate receptors. GPCRs are excellent drug targets in various neurologic and psychiatric diseases. All antipsychotics have been shown to act at the dopamine D2 receptor; similarly novel antipsychotics also act at the serotonergic receptor, and often the muscarinic and adrenergic receptors as well. While the majority of antidepressants can be classified as selective serotonin reuptake inhibitors, blockade of the 5-HT1A and α_2 adrenergic receptors increases the effects of these drugs. The GPCRs are also of use as drug targets in the treatment of stroke. Blockade of the glutamate receptors may decrease the neuronal death resulting from excitotoxicity; further more the purinergic receptors have also been implicated as drug targets in the treatment of cerebral ischemia. The β -adrenergic receptors have been implicated in the treatment of ADHD with Ritalin, while the α -adrenergic receptors have been implicated in memory. Therefore this gene may be of use as a small molecule target for the treatment of any of the described diseases.

In addition, this gene is expressed in clusters of cell lines derived from lung cancer and colon cancer. Thus, expression of this gene could be used to differentiate between these sample and other samples on this panel and as a marker to detect the presence of colon and lung cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung and colon cancers.

References:

El Yacoubi M, Ledent C, Parmentier M, Bertorelli R, Ongini E, Costentin J, Vaugeois JM. Adenosine A2A receptor antagonists are potential antidepressants: evidence based on pharmacology and A2A receptor knockout mice. Br J Pharmacol 2001 Sep;134(1):68-77

1. Adenosine, an ubiquitous neuromodulator, and its analogues have been shown to produce 'depressant' effects in animal models believed to be relevant to depressive disorders, while adenosine receptor antagonists have been found to reverse adenosine-mediated 'depressant' effect. 2. We have designed studies to assess whether adenosine A2A receptor antagonists, or genetic inactivation of the receptor would be effective in established screening procedures, such as tail suspension and forced swim tests, which are predictive of clinical antidepressant activity. 3. Adenosine A2A receptor knockout mice were found to be less sensitive to 'depressant' challenges than their wildtype littermates. Consistently, the adenosine A2A receptor blockers SCH 58261 (1 - 10 mg kg⁻¹, i.p.) and KW 6002 (0.1 - 10 mg kg⁻¹, p.o.) reduced the total immobility time in the tail suspension test. 4. The efficacy of adenosine A2A receptor antagonists in reducing immobility time in the tail suspension test was confirmed and extended in two groups of mice. Specifically, SCH 58261 (1 - 10 mg kg⁻¹) and ZM 241385 (15 - 60 mg kg⁻¹) were effective in mice previously screened for having high immobility time, while SCH 58261 at 10 mg kg⁻¹ reduced immobility of mice that were selectively bred for their spontaneous 'helplessness' in this assay. 5. Additional experiments were carried out using the forced swim test. SCH 58261 at 10 mg kg⁻¹ reduced the immobility time by 61%, while KW 6002 decreased the total immobility time at the doses of 1 and 10 mg kg⁻¹ by 75 and 79%, respectively. 6. Administration of the dopamine D2 receptor antagonist haloperidol (50 - 200 microg kg⁻¹ i.p.) prevented the antidepressant-like effects elicited by SCH 58261 (10 mg kg⁻¹ i.p.) in forced swim test whereas it left unaltered its stimulant motor effects. 7. In conclusion, these data support the hypothesis that A2A receptor antagonists prolong escape-directed behaviour in two screening tests for antidepressants. Altogether the results support the hypothesis that blockade of the adenosine A2A receptor might be an interesting target for the development of effective antidepressant agents.

25 Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. Clin Psychiatry 2001;62 Suppl 15:12-7

Although selective serotonin reuptake inhibitors (SSRIs) block serotonin (5-HT) reuptake rapidly, their therapeutic action is delayed. The increase in synaptic 5-HT activates feedback mechanisms mediated by 5-HT_{1A} (cell body) and 5-HT_{1B} (terminal) autoreceptors, which, respectively, reduce the firing in 5-HT neurons and decrease the amount of 5-HT released per action potential resulting in attenuated 5-HT neurotransmission. Long-term treatment desensitizes the inhibitory 5-HT₁ autoreceptors, and 5-HT neurotransmission is enhanced. The time course of these events is similar to the delay of clinical action. The addition of pindolol, which blocks 5-HT_{1A} receptors, to SSRI treatment decouples the

feedback inhibition of 5-HT neuron firing and accelerates and enhances the antidepressant response. The neuronal circuitry of the 5-HT and norepinephrine (NE) systems and their connections to forebrain areas believed to be involved in depression has been dissected. The firing of 5-HT neurons in the raphe nuclei is driven, at least partly, by alpha1-adrenoceptor-mediated excitatory inputs from NE neurons. Inhibitory alpha2-adrenoceptors on the NE neuroterminals form part of a feedback control mechanism. Mirtazapine, an antagonist at alpha2-adrenoceptors, does not enhance 5-HT neurotransmission directly but disinhibits the NE activation of 5-HT neurons and thereby increases 5-HT neurotransmission by a mechanism that does not require a time-dependent desensitization of receptors. These neurobiological phenomena may underlie the apparently faster onset of action of mirtazapine compared with the SSRIs.

Tranquillini ME, Reggiani A. Glycine-site antagonists and stroke. *Expert Opin Investig Drugs* 1999 Nov;8(11):1837-1848

The excitatory amino acid, (S)-glutamic acid, plays an important role in controlling many neuronal processes. Its action is mediated by two main groups of receptors: the ionotropic receptors (which include NMDA, AMPA and kainic acid subtypes) and the metabotropic receptors (mGluR(1-8)) mediating G-protein coupled responses. This review focuses on the strychnine insensitive glycine binding site located on the NMDA receptor channel, and on the possible use of selective antagonists for the treatment of stroke. Stroke is a devastating disease caused by a sudden vascular accident. Neurochemically, a massive release of glutamate occurs in neuronal tissue; this overactivates the NMDA receptor, leading to increased intracellular calcium influx, which causes neuronal cell death through necrosis. NMDA receptor activation strongly depends upon the presence of glycine as a co-agonist. Therefore, the administration of a glycine antagonist can block overactivation of NMDA receptors, thus preserving neurones from damage. The glycine antagonists currently identified can be divided into five main categories depending on their chemical structure: indoles, tetrahydroquinolines, benzoazepines, quinoxalinediones and pyrida-zinoquinolines.

Monopoli A, Lozza G, Forlani A, Mattavelli A, Ongini E. Blockade of adenosine A2A receptors by SCH 58261 results in neuroprotective effects in cerebral ischaemia in rats.

Neuroreport 1998 Dec 1;9(17):3955-9

Blockade of adenosine receptors can reduce cerebral infarct size in the model of global ischaemia. Using the potent and selective A2A adenosine receptor antagonist, SCH 58261, we assessed whether A2A receptors are involved in the neuronal damage following focal cerebral ischaemia as induced by occluding the left middle cerebral artery. SCH 58261 (0.01 mg/kg

either i.p. or i.v.) administered to normotensive rats 10 min after ischaemia markedly reduced cortical infarct volume as measured 24 h later (30% vs controls, $p < 0.05$). Similar effects were observed when SCH 58261 (0.01 mg/kg, i.p.) was administered to hypertensive rats (28% infarct volume reduction vs controls, $p < 0.05$). Neuroprotective properties of SCH 58261 administered after ischaemia indicate that blockade of A2A adenosine receptors is a potentially useful biological target for the reduction of brain injury.

Panel 2D Summary: Ag2913 The NOV66 gene is a diagnostic marker for gastric thyroid and bladder cancer and a target for therapeutic intervention in gastric, thyroid and bladder cancer through the use of antibodies or small molecule drugs. This is based on the expression profile of this gene that shows higher expression in some gastric, thyroid and bladder cancer samples compared to normal tissues.

Panel 4D Summary: Ag2913 The NOV66 gene, an olfactory receptor homolog is expressed at moderate levels in activated and resting T lymphocytes (CT range 30.13-32.98). Small molecules or therapeutic antibodies that antagonize the function of the NOV66 gene product may reduce or eliminate the symptoms of autoimmune and inflammatory diseases, including Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

Panel CNS_1 Summary: Ag2913 No significant expression detected. Potential probe/primer failure (data not shown).

NOV67

Expression of gene NOV67 was assessed using the primer-probe set Ag2951, described in Table BJA. Results of the RTQ-PCR runs are shown in Tables BJB and BJC.

Table BJA. Probe Name Ag2951

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-acctctcacttggtggacatctg-3'	22	243	1270
Probe	TET-5'-tacacctccagcaggggtccctcagat-3'-TAMRA	26	266	1271
Reverse	5'-ggcaaaggagatggtctttct-3'	21	314	1272

Table BJB. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag2951, Run 248065290	Tissue Name	Rel. Exp.(%) Ag2951, Run 248065290
110967 COPD-F	3.1	112427 Match Control Psoriasis-F	49.3

110980 COPD-F	4.9	112418 Psoriasis-M	13.6
110968 COPD-M	2.4	112723 Match Control Psoriasis-M	0.0
110977 COPD-M	45.7	112419 Psoriasis-M	20.2
110989 Emphysema-F	25.9	112424 Match Control Psoriasis-M	9.5
110992 Emphysema-F	33.9	112420 Psoriasis-M	93.3
110993 Emphysema-F	5.8	112425 Match Control Psoriasis-M	32.8
110994 Emphysema-F	0.0	104689 (MF) OA Bone-Backus	0.0
110995 Emphysema-F	55.5	104690 (MF) Adj "Normal" Bone-Backus	0.0
110996 Emphysema-F	7.0	104691 (MF) OA Synovium-Backus	0.0
110997 Asthma-M	12.8	104692 (BA) OA Cartilage-Backus	0.0
111001 Asthma-F	18.6	104694 (BA) OA Bone-Backus	0.0
111002 Asthma-F	27.7	104695 (BA) Adj "Normal" Bone-Backus	5.5
111003 Atopic Asthma-F	16.7	104696 (BA) OA Synovium-Backus	10.2
111004 Atopic Asthma-F	37.4	104700 (SS) OA Bone-Backus	7.1
111005 Atopic Asthma-F	31.6	104701 (SS) Adj "Normal" Bone-Backus	14.3
111006 Atopic Asthma-F	1.8	104702 (SS) OA Synovium-Backus	28.3
111417 Allergy-M	20.4	117093 OA Cartilage Rep7	65.5
112347 Allergy-M	2.5	112672 OA Bone5	26.6
112349 Normal Lung-F	0.0	112673 OA Synovium5	21.6
112357 Normal Lung-F	12.8	112674 OA Synovial Fluid cells5	9.3
112354 Normal Lung-M	4.5	117100 OA Cartilage Rep14	2.9
112374 Crohns-F	16.3	112756 OA Bone9	7.3
112389 Match Control Crohns-F	15.0	112757 OA Synovium9	3.3
112375 Crohns-F	7.5	112758 OA Synovial Fluid Cells9	12.9
112732 Match Control Crohns-F	41.5	117125 RA Cartilage Rep2	0.0

112725 Crohns-M	2.0	113492 Bone2 RA	8.5
112387 Match Control Crohns-M	5.3	113493 Synovium2 RA	1.9
112378 Crohns-M	4.0	113494 Syn Fluid Cells RA	7.6
112390 Match Control Crohns-M	61.6	113499 Cartilage4 RA	13.7
112726 Crohns-M	21.3	113500 Bone4 RA	8.3
112731 Match Control Crohns-M	14.6	113501 Synovium4 RA	20.4
112380 Ulcer Col-F	18.3	113502 Syn Fluid Cells4 RA	6.5
112734 Match Control Ulcer Col-F	100.0	113495 Cartilage3 RA	11.7
112384 Ulcer Col-F	60.7	113496 Bone3 RA	4.4
112737 Match Control Ulcer Col-F	22.1	113497 Synovium3 RA	3.0
112386 Ulcer Col-F	4.6	113498 Syn Fluid Cells3 RA	17.3
112738 Match Control Ulcer Col-F	3.7	117106 Normal Cartilage Rep20	11.4
112381 Ulcer Col-M	0.0	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	15.4	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	58.2	113665 Syn Fluid Cells3 Normal	1.3
112394 Match Control Ulcer Col-M	4.2	117107 Normal Cartilage Rep22	0.0
112383 Ulcer Col-M	63.3	113667 Bone4 Normal	39.5
112736 Match Control Ulcer Col-M	12.1	113668 Synovium4 Normal	25.5
112423 Psoriasis-F	26.2	113669 Syn Fluid Cells4 Normal	26.1

Table BJC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2951, Run 164403342	Tissue Name	Rel. Exp.(%) Ag2951, Run 164403342
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.8
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0

Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.7	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	1.3	Bronchial epithelium TNFalpha + IL1beta	1.0
Primary Th2 rest	7.0	Small airway epithelium none	0.0
Primary Tr1 rest	17.2	Small airway epithelium TNFalpha + IL-1beta	5.2
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	1.5	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	1.0
CD4 lymphocyte none	13.3	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 _anti- CD95 CH11	0.9	CCD1106 (Keratinocytes) none	1.2
LAK cells rest	11.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	7.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	3.2	NCI-H292 none	2.4
LAK cells IL-2+ IL-18	4.6	NCI-H292 IL-4	1.9
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	1.6	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	7.6	Lung fibroblast none	0.0
PBMC PWM	100.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	5.8	Lung fibroblast IL-4	0.0

Ramos (B cell) none	0.0	Lung fibroblast IL-9	1.7
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	1.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	5.2	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	19.9	Dermal fibroblast CCD1070 TNF alpha	1.4
EOL-1 dbcAMP PMA/ionomycin	7.8	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	9.2	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	2.9	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	3.0	IBD Colitis 2	1.8
Monocytes rest	26.8	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	2.5
Macrophages rest	2.4	Lung	1.2
Macrophages LPS	0.0	Thymus	0.8
HUVEC none	0.0	Kidney	8.7
HUVEC starved	0.0		

AI_comprehensive_panel_v1.0 Summary: Ag2951 Highest expression of the NOV67 gene is seen in normal tissue adjacent to colon from an ulcerative colitis patient (CT=33).

Thus, expression of this gene could be used to distinguish this sample from other samples on this panel. Please see Panel 4D for further discussion of utility of this gene in inflammation.

CNS_neurodegeneration_v1.0 Summary: Ag2951 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag2951 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag2951 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag2951 The NOV67 gene is expressed at a moderate level (CT=32.78) in pokeweed mitogen-stimulated peripheral blood leukocytes, consisting primarily of activated B lymphocytes. Small molecule antagonists or therapeutic antibody antagonists that block the function of the CG56571-gene product may be useful in several autoimmune and inflammatory diseases in which activated B cells can play major roles as sources of autoantibody-producing cells and as powerful antigen-presenting cells, including, but not

limited to, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

NOV69a and NOV69b

Expression of gene NOV69a and variant NOV69b was assessed using the primer-probe sets Ag2460 and Ag349, described in Tables BKA and BKB. Results of the RTQ-PCR runs are shown in Tables BKC, BKD, BKE and BKF.

Table BKA. Probe Name Ag2460

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tcatagcagtcctccgaggaa-3'	19	89	1273
Probe	TET-5'-tcactattgccttaatctcatgccga-3'-TAMRA	26	125	1274
Reverse	5'-ttctcaagggtctccacatg-3'	20	151	1275

Table BKB. Probe Name Ag349

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gggaaagccacagactcgaa-3'	20	289	1276
Probe	TET-5'-cttctaccacagccagagtggcaggaact-3'-TAMRA	29	255	1277
Reverse	5'-acccgagcctgtgaagtcct-3'	20	231	1278

10

Table BKC. Panel 1

Tissue Name	Rel. Exp.(%) Ag349, Run 97804233	Tissue Name	Rel. Exp.(%) Ag349, Run 97804233
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	0.2
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.0
Adrenal gland	0.0	Renal ca. UO-31	0.0
Thyroid	0.5	Renal ca. TK-10	0.0
Salivary gland	25.5	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell)	0.0

		LX-1	
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	1.5
Brain (substantia nigra)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Brain (thalamus)	0.0	Lung ca. (large cell)NCI-H460	1.3
Brain (hypothalamus)	0.0	Lung ca. (non-sm. cell) A549	0.0
Spinal cord	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
glio/astro U87-MG	59.0	Lung ca. (non-s.cell) HOP-62	0.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SW1783	0.0	Lung ca. (squam.) SW 900	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (squam.) NCI-H596	0.0
astrocytoma SF-539	0.0	Mammary gland	0.0
astrocytoma SNB-75	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SNB-19	3.1	Breast ca.* (pl.ef) MDA-MB-231	0.0
glioma U251	0.0	Breast ca.* (pl. ef) T47D	0.0
glioma SF-295	0.0	Breast ca. BT-549	0.0
Heart	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.0
Thymus	66.4	Ovarian ca. OVCAR- 4	0.0
Spleen	0.1	Ovarian ca. OVCAR- 5	1.5
Lymph node	3.1	Ovarian ca. OVCAR- 8	0.0
Colon (ascending)	29.9	Ovarian ca. IGROV-1	0.0
Stomach	77.4	Ovarian ca. (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620 (SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone	0.0

		met) PC-3	
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. HCT-15	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca. * (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	15.5	Melanoma* (met) SK- MEL-5	0.0
Kidney	0.0	Melanoma SK-MEL- 28	100.0
Kidney (fetal)	0.0		

Table BKD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2460, Run 157914666	Tissue Name	Rel. Exp.(%) Ag2460, Run 157914666
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	9.6
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	5.6	Renal ca. ACHN	0.0
Salivary gland	19.8	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	2.2
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	45.7	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	2.4	Lung ca. (non-sm. cell) A549	0.0

glio/astro U-118-MG	2.5	Lung ca. (non-s.cell) NCI-H23	0.0
astrocytoma SW1783	0.0	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	0.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	100.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR- 3	0.0
Thymus	61.1	Ovarian ca. OVCAR- 4	0.0
Spleen	0.0	Ovarian ca. OVCAR- 5	0.0
Lymph node	8.4	Ovarian ca. OVCAR- 8	0.0
Colorectal	0.0	Ovarian ca. IGROV- 1	0.0
Stomach	77.4	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	2.3
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met)	0.0	Melanoma M14	0.0

NCI-N87			
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	59.5	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table BKE. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2460, Run 157914720	Tissue Name	Rel. Exp.(%) Ag2460, Run 157914720
Normal Colon	0.0	Kidney Margin 8120608	0.0
CC Well to Mod Diff (ODO3866)	0.0	Kidney Cancer 8120613	0.0
CC Margin (ODO3866)	0.0	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer 9010320	0.0
CC Margin (ODO3868)	0.0	Kidney Margin 9010321	0.0
CC Mod Diff (ODO3920)	0.0	Normal Uterus	0.0
CC Margin (ODO3920)	0.0	Uterus Cancer 064011	0.0
CC Gr.2 ascend colon (ODO3921)	0.0	Normal Thyroid	0.9
CC Margin (ODO3921)	0.0	Thyroid Cancer 064010	0.0
CC from Partial Hepatectomy (ODO4309) Mets	0.0	Thyroid Cancer A302152	0.0
Liver Margin (ODO4309)	0.0	Thyroid Margin A302153	0.0
Colon mets to lung (OD04451-01)	0.0	Normal Breast	0.0
Lung Margin (OD04451- 02)	0.0	Breast Cancer (OD04566)	0.0
Normal Prostate 6546-1	0.0	Breast Cancer (OD04590-01)	0.0
Prostate Cancer (OD04410)	0.0	Breast Cancer Mets (OD04590-03)	0.2
Prostate Margin (OD04410)	0.0	Breast Cancer Metastasis (OD04655-05)	0.6
Prostate Cancer (OD04720-01)	0.0	Breast Cancer 064006	0.0

Prostate Margin (OD04720-02)	0.0	Breast Cancer 1024	0.0
Normal Lung 061010	0.4	Breast Cancer 9100266	0.0
Lung Met to Muscle (ODO4286)	0.0	Breast Margin 9100265	0.0
Muscle Margin (ODO4286)	0.0	Breast Cancer A209073	0.0
Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	100.0	Liver Cancer 064003	0.0
Lung Margin (OD04404)	0.0	Liver Cancer 1025	0.1
Lung Cancer (OD04565)	0.5	Liver Cancer 1026	0.0
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237- 01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237- 02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	0.0
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	Bladder Normal Adjacent (OD04718- 03)	0.0
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	0.0	Normal Stomach	0.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	0.0

Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	0.0	Stomach Margin 9060396	0.0
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	0.6

Table BKF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2460, Run 157914794	Tissue Name	Rel. Exp.(%) Ag2460, Run 157914794
Secondary Th1 act	0.4	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.5	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	2.3	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.5	HUVEC IL-11	0.0
Secondary Tr1 rest	3.7	Lung Microvascular EC none	0.0
Primary Th1 act	0.6	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	1.1	Microvascular Dermal EC none	0.0
Primary Tr1 act	1.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	15.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	9.2	Small airway epithelium none	2.1
Primary Tr1 rest	4.3	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	0.9	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	1.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	1.1	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.5	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.5	KU-812 (Basophil) rest	0.0

CD4 lymphocyte none	1.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 anti- CD95 CH11	2.6	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.5
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.5	NCI-H292 none	0.3
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.6
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	1.1
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.8	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	2.1	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	2.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	1.8	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	1.6	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.8
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti- CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	1.1	Colon	0.0
Macrophages rest	0.6	Lung	0.0
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	9.9
HUVEC starved	0.0		

References:

Smith,D.E., Renshaw,B.R., Ketchem,R.R., Kubin,M., Garka,K.E. and Sims,J.E.Four new members expand the interleukin-1 superfamily J. Biol. Chem. 275 (2), 1169-1175 (2000)

Abstract: We report here the cloning and characterization of four new members of the interleukin-1 (IL-1) family (FIL1delta, FIL1epsilon, FIL1zeta, and FIL1eta, with FIL1 standing for "Family of IL-1"). The novel genes demonstrate significant sequence similarity to IL-1alpha, IL-1beta, IL-1ra, and IL-18, and in addition maintain a conserved exon-intron arrangement that is shared with the previously known members of the family. Protein structure modeling also suggests that the FIL1 genes are related to IL-1beta and IL-1ra. The novel genes form a cluster with the IL-1s on the long arm of human chromosome 2.

NOV71

Expression of gene NOV71 was assessed using the primer-probe set Ag3049, described in Table BLA. Results of the RTQ-PCR runs are shown in Tables BLB, BLC, BLD and BLE.

Table BLA. Probe Name Ag3049

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gctggcgatatattaatgaattg-3'	22	696	1279
Probe	TET-5'-catgcagacgtggatctttacgcact-3'-TAMRA	26	718	1280
Reverse	5'-agtacaatggcaacagcatcat-3'	22	767	1281

Table BLB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3049, Run 209823735	Tissue Name	Rel. Exp.(%) Ag3049, Run 209823735
AD 1 Hippo	10.2	Control (Path) 3 Temporal Ctx	5.9
AD 2 Hippo	33.7	Control (Path) 4 Temporal Ctx	28.1
AD 3 Hippo	9.3	AD 1 Occipital Ctx	14.2
AD 4 Hippo	7.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	99.3	AD 3 Occipital Ctx	5.8
AD 6 Hippo	46.3	AD 4 Occipital Ctx	16.0
Control 2 Hippo	44.1	AD 5 Occipital Ctx	17.1
Control 4 Hippo	6.0	AD 6 Occipital Ctx	63.7
Control (Path) 3 Hippo	7.6	Control 1 Occipital Ctx	2.2

AD 1 Temporal Ctx	13.0	Control 2 Occipital Ctx	68.3
AD 2 Temporal Ctx	33.4	Control 3 Occipital Ctx	11.5
AD 3 Temporal Ctx	6.8	Control 4 Occipital Ctx	5.5
AD 4 Temporal Ctx	16.5	Control (Path) 1 Occipital Ctx	77.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	8.6
AD 5 Sup Temporal Ctx	48.0	Control (Path) 3 Occipital Ctx	2.1
AD 6 Inf Temporal Ctx	37.9	Control (Path) 4 Occipital Ctx	12.9
AD 6 Sup Temporal Ctx	41.5	Control 1 Parietal Ctx	4.1
Control 1 Temporal Ctx	6.3	Control 2 Parietal Ctx	29.9
Control 2 Temporal Ctx	62.4	Control 3 Parietal Ctx	18.0
Control 3 Temporal Ctx	14.2	Control (Path) 1 Parietal Ctx	97.9
Control 4 Temporal Ctx	7.7	Control (Path) 2 Parietal Ctx	19.3
Control (Path) 1 Temporal Ctx	64.6	Control (Path) 3 Parietal Ctx	6.3
Control (Path) 2 Temporal Ctx	29.9	Control (Path) 4 Parietal Ctx	42.6

Table BLC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3049, Run 167972763	Tissue Name	Rel. Exp.(%) Ag3049, Run 167972763
Liver adenocarcinoma	37.6	Kidney (fetal)	7.5
Pancreas	1.8	Renal ca. 786-0	4.7
Pancreatic ca. CAPAN 2	8.0	Renal ca. A498	8.7
Adrenal gland	2.1	Renal ca. RXF 393	28.3
Thyroid	3.0	Renal ca. ACHN	6.3
Salivary gland	1.5	Renal ca. UO-31	36.9
Pituitary gland	8.5	Renal ca. TK-10	9.5
Brain (fetal)	10.4	Liver	0.2
Brain (whole)	9.0	Liver (fetal)	1.7
Brain (amygdala)	18.7	Liver ca. (hepatoblast) HepG2	6.3

Brain (cerebellum)	32.5	Lung	0.4
Brain (hippocampus)	14.0	Lung (fetal)	2.0
Brain (substantia nigra)	6.0	Lung ca. (small cell) LX-1	5.4
Brain (thalamus)	2.9	Lung ca. (small cell) NCI-H69	3.6
Cerebral Cortex	30.1	Lung ca. (s.cell var.) SHP-77	15.9
Spinal cord	3.4	Lung ca. (large cell)NCI-H460	1.7
glio/astro U87-MG	27.4	Lung ca. (non-sm. cell) A549	28.9
glio/astro U-118-MG	16.0	Lung ca. (non-s.cell) NCI-H23	19.6
astrocytoma SW1783	40.3	Lung ca. (non-s.cell) HOP-62	18.3
neuro*; met SK-N-AS	0.6	Lung ca. (non-s.cl) NCI-H522	23.5
astrocytoma SF-539	2.9	Lung ca. (squam.) SW 900	15.4
astrocytoma SNB-75	26.8	Lung ca. (squam.) NCI-H596	4.0
glioma SNB-19	34.4	Mammary gland	1.1
glioma U251	100.0	Breast ca.* (pl.ef) MCF-7	37.4
glioma SF-295	39.2	Breast ca.* (pl.ef) MDA-MB-231	25.7
Heart (fetal)	1.3	Breast ca.* (pl.ef) T47D	55.5
Heart	1.2	Breast ca. BT-549	7.4
Skeletal muscle (fetal)	2.0	Breast ca. MDA-N	2.0
Skeletal muscle	9.3	Ovary	2.7
Bone marrow	0.3	Ovarian ca. OVCAR- 3	2.6
Thymus	3.0	Ovarian ca. OVCAR- 4	17.4
Spleen	2.2	Ovarian ca. OVCAR- 5	50.7
Lymph node	5.0	Ovarian ca. OVCAR- 8	1.1
Colorectal	1.9	Ovarian ca. IGROV- 1	5.2
Stomach	3.3	Ovarian ca.* (ascites) SK-OV-3	24.7
Small intestine	0.2	Uterus	0.1

Colon ca. SW480	7.9	Placenta	0.2
Colon ca.* SW620(SW480 met)	30.4	Prostate	3.7
Colon ca. HT29	8.7	Prostate ca.* (bone met)PC-3	91.4
Colon ca. HCT-116	12.4	Testis	0.5
Colon ca. CaCo-2	18.7	Melanoma Hs688(A).T	6.0
Colon ca. tissue(ODO3866)	3.7	Melanoma* (met) Hs688(B).T	10.1
Colon ca. HCC-2998	8.9	Melanoma UACC-62	11.0
Gastric ca.* (liver met) NCI-N87	8.7	Melanoma M14	0.4
Bladder	10.7	Melanoma LOX IMVI	17.8
Trachea	1.0	Melanoma* (met) SK-MEL-5	1.7
Kidney	2.2	Adipose	3.9

Table BLD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3049, Run 174441445	Tissue Name	Rel. Exp.(%) Ag3049, Run 174441445
Normal Colon	11.7	Kidney Margin (OD04348)	64.6
Colon cancer (OD06064)	57.4	Kidney malignant cancer (OD06204B)	10.5
Colon Margin (OD06064)	3.4	Kidney normal adjacent tissue (OD06204E)	10.1
Colon cancer (OD06159)	12.1	Kidney Cancer (OD04450-01)	24.1
Colon Margin (OD06159)	1.8	Kidney Margin (OD04450-03)	21.6
Colon cancer (OD06297-04)	7.1	Kidney Cancer 8120613	9.9
Colon Margin (OD06297-015)	14.7	Kidney Margin 8120614	13.6
CC Gr.2 ascend colon (ODO3921)	5.1	Kidney Cancer 9010320	10.3
CC Margin (ODO3921)	5.4	Kidney Margin 9010321	11.3
Colon cancer metastasis (OD06104)	1.8	Kidney Cancer 8120607	30.6
Lung Margin (OD06104)	14.8	Kidney Margin 8120608	11.2

Colon mets to lung (OD04451-01)	25.5	Normal Uterus	2.0
Lung Margin (OD04451-02)	51.4	Uterine Cancer 064011	6.8
Normal Prostate	28.9	Normal Thyroid	6.3
Prostate Cancer (OD04410)	13.4	Thyroid Cancer 064010	14.2
Prostate Margin (OD04410)	20.3	Thyroid Cancer A302152	51.4
Normal Ovary	7.9	Thyroid Margin A302153	8.5
Ovarian cancer (OD06283-03)	56.3	Normal Breast	53.6
Ovarian Margin (OD06283-07)	13.6	Breast Cancer (OD04566)	6.9
Ovarian Cancer 064008	24.1	Breast Cancer 1024	100.0
Ovarian cancer (OD06145)	31.2	Breast Cancer (OD04590-01)	9.5
Ovarian Margin (OD06145)	14.0	Breast Cancer Mets (OD04590-03)	24.3
Ovarian cancer (OD06455-03)	10.4	Breast Cancer Metastasis (OD04655-05)	65.1
Ovarian Margin (OD06455-07)	1.0	Breast Cancer 064006	26.2
Normal Lung	11.8	Breast Cancer 9100266	24.5
Invasive poor diff. lung adeno (ODO4945-01)	10.7	Breast Margin 9100265	25.0
Lung Margin (ODO4945-03)	6.2	Breast Cancer A209073	19.6
Lung Malignant Cancer (OD03126)	5.9	Breast Margin A2090734	58.6
Lung Margin (OD03126)	3.3	Breast cancer (OD06083)	79.0
Lung Cancer (OD05014A)	42.3	Breast cancer node metastasis (OD06083)	40.1
Lung Margin (OD05014B)	13.0	Normal Liver	31.0
Lung cancer (OD06081)	13.1	Liver Cancer 1026	12.5
Lung Margin (OD06081)	8.1	Liver Cancer 1025	25.5
Lung Cancer (OD04237-01)	35.8	Liver Cancer 6004-T	14.8
Lung Margin (OD04237-02)	14.7	Liver Tissue 6004-N	4.7
Ocular Melanoma	0.0	Liver Cancer 6005-T	25.9

Metastasis			
Ocular Melanoma Margin (Liver)	8.5	Liver Tissue 6005-N	70.7
Melanoma Metastasis	4.3	Liver Cancer 064003	14.5
Melanoma Margin (Lung)	6.5	Normal Bladder	32.3
Normal Kidney	16.5	Bladder Cancer 1023	6.6
Kidney Ca, Nuclear grade 2 (OD04338)	41.2	Bladder Cancer A302173	51.1
Kidney Margin (OD04338)	10.8	Normal Stomach	72.7
Kidney Ca Nuclear grade 1/2 (OD04339)	45.7	Gastric Cancer 9060397	4.7
Kidney Margin (OD04339)	8.2	Stomach Margin 9060396	40.6
Kidney Ca, Clear cell type (OD04340)	18.4	Gastric Cancer 9060395	10.4
Kidney Margin (OD04340)	7.4	Stomach Margin 9060394	57.8
Kidney Ca, Nuclear grade 3 (OD04348)	17.9	Gastric Cancer 064005	24.7

Table BLE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3049, Run 164334396	Tissue Name	Rel. Exp.(%) Ag3049, Run 164334396
Secondary Th1 act	9.4	HUVEC IL-1beta	1.0
Secondary Th2 act	7.8	HUVEC IFN gamma	3.1
Secondary Tr1 act	8.1	HUVEC TNF alpha + IFN gamma	2.1
Secondary Th1 rest	0.3	HUVEC TNF alpha + IL4	1.0
Secondary Th2 rest	2.0	HUVEC IL-11	0.7
Secondary Tr1 rest	2.1	Lung Microvascular EC none	0.3
Primary Th1 act	13.0	Lung Microvascular EC TNFalpha + IL-1beta	0.2
Primary Th2 act	13.5	Microvascular Dermal EC none	0.5
Primary Tr1 act	16.2	Microsvascular Dermal EC TNFalpha + IL-1beta	0.1
Primary Th1 rest	3.8	Bronchial epithelium TNFalpha + IL1beta	4.5
Primary Th2 rest	3.7	Small airway epithelium none	2.6

Primary Tr1 rest	1.0	Small airway epithelium TNFalpha + IL-1beta	8.5
CD45RA CD4 lymphocyte act	6.6	Coronary artery SMC rest	11.8
CD45RO CD4 lymphocyte act	7.8	Coronary artery SMC TNFalpha + IL-1beta	6.8
CD8 lymphocyte act	5.6	Astrocytes rest	3.8
Secondary CD8 lymphocyte rest	4.0	Astrocytes TNFalpha + IL-1beta	5.4
Secondary CD8 lymphocyte act	9.2	KU-812 (Basophil) rest	1.5
CD4 lymphocyte none	0.6	KU-812 (Basophil) PMA/ionomycin	10.4
2ry Th1/Th2/Tr1_anti- CD95 CH11	3.7	CCD1106 (Keratinocytes) none	3.7
LAK cells rest	9.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	4.5
LAK cells IL-2	3.0	Liver cirrhosis	0.5
LAK cells IL-2+IL-12	4.6	Lupus kidney	0.3
LAK cells IL-2+IFN gamma	9.6	NCI-H292 none	9.5
LAK cells IL-2+ IL-18	9.5	NCI-H292 IL-4	15.1
LAK cells PMA/ionomycin	4.7	NCI-H292 IL-9	12.5
NK Cells IL-2 rest	1.4	NCI-H292 IL-13	7.7
Two Way MLR 3 day	3.6	NCI-H292 IFN gamma	9.0
Two Way MLR 5 day	3.7	HPAEC none	2.1
Two Way MLR 7 day	2.3	HPAEC TNF alpha + IL-1 beta	1.1
PBMC rest	1.1	Lung fibroblast none	2.5
PBMC PWM	17.6	Lung fibroblast TNF alpha + IL-1 beta	1.2
PBMC PHA-L	10.8	Lung fibroblast IL-4	5.0
Ramos (B cell) none	23.2	Lung fibroblast IL-9	7.5
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	3.9
B lymphocytes PWM	40.9	Lung fibroblast IFN gamma	14.6
B lymphocytes CD40L and IL-4	25.2	Dermal fibroblast CCD1070 rest	9.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	14.8
EOL-1 dbcAMP PMA/ionomycin	0.4	Dermal fibroblast CCD1070 IL-1 beta	6.0
Dendritic cells none	7.7	Dermal fibroblast IFN	10.2



5

10

15

20

References:

25

Little is known of the functional properties of the mammalian, brain-specific Na(+)/H(+) exchanger isoform 5 (NHE5). Rat NHE5 was stably expressed in NHE-deficient PS120 cells, and its activity was characterized using the fluorescent pH-sensitive dye 2',7'-bis(2-carboxyethyl)-5(6)-carboxyfluorescein. NHE5 was insensitive to ethylisopropyl
5 amiloride. The transport kinetics displayed a simple Michaelis-Menten relationship for extracellular Na(+) (apparent K(Na) = 27 +/- 5 mM) and a Hill coefficient near 3 for the intracellular proton concentration with a half-maximal activity at an intracellular pH of 6.93 +/- 0.03. NHE5 activity was inhibited by acute exposure to 8-bromo-cAMP or forskolin (which increases intracellular cAMP by activating adenylate cyclase). The kinase inhibitor H-
10 89 reversed this inhibition, suggesting that regulation by cAMP involves a protein kinase A (PKA)-dependent process. In contrast, 8-bromo-cGMP did not have a significant effect on activity. The protein kinase C (PKC) activator phorbol 12-myristate 13-acetate inhibited NHE5, and the PKC antagonist chelerythrine chloride blunted this effect. Activity was also inhibited by hyperosmotic-induced cell shrinkage but was unaffected by a hyposmotic
15 challenge. These results demonstrate that rat brain NHE5 is downregulated by activation of PKA and PKC and by cell shrinkage, important regulators of neuronal cell function.

Panel 2.2 Summary: Ag3049 The NOV71 gene can be used as a diagnostic marker for stomach, breast, lung, ovarian and some colon cancers as expression in the normal adjacent tissue and the tumor tissue differs. Antibodies and small molecule inhibitors designed with this
20 gene product may also be used for therapy in breast, lung, ovarian and some colon cancers.

Panel 4D Summary: Ag3049 The NOV71 gene, a sodium/hydrogen Exchanger homolog is expressed at a high level in Ramos (B cell) activated with ionomycin (CT=24.72), and at a moderate to high level in other activated B cell preparations. Therefore, small molecule antagonists or therapeutic antibody antagonists that block the function of the NOV71
25 gene product may be useful in several autoimmune and inflammatory diseases in which activated B cells can play major roles as sources of autoantibody-producing cells and also as powerful antigen-presenting cells, including, but not limited to, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

30 **NOV72: UBIQUITIN-SPECIFIC PROTEASE**

Expression of gene NOV72 was assessed using the primer-probe set Ag3050, described in Table BMA. Results of the RTQ-PCR runs are shown in Tables BMB, BMC, BMD and BME.

Table BMA. Probe Name Ag3050

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgaccaggtattaacctggaa-3'	22	799	1282
Probe	TET-5'-ttactgctgcagggacatgctctcct-3'-TAMRA	26	823	1283
Reverse	5'-taggcaaaggtctctttgtcaa-3'	22	852	1284

Table BMB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3050, Run 211012446	Tissue Name	Rel. Exp.(%) Ag3050, Run 211012446
AD 1 Hippo	12.5	Control (Path) 3 Temporal Ctx	4.3
AD 2 Hippo	22.1	Control (Path) 4 Temporal Ctx	29.1
AD 3 Hippo	3.0	AD 1 Occipital Ctx	15.0
AD 4 Hippo	11.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	59.0	AD 3 Occipital Ctx	7.1
AD 6 Hippo	51.1	AD 4 Occipital Ctx	22.2
Control 2 Hippo	17.6	AD 5 Occipital Ctx	24.7
Control 4 Hippo	6.1	AD 6 Occipital Ctx	22.4
Control (Path) 3 Hippo	7.0	Control 1 Occipital Ctx	2.9
AD 1 Temporal Ctx	17.4	Control 2 Occipital Ctx	39.0
AD 2 Temporal Ctx	23.5	Control 3 Occipital Ctx	16.8
AD 3 Temporal Ctx	7.9	Control 4 Occipital Ctx	6.0
AD 4 Temporal Ctx	27.5	Control (Path) 1 Occipital Ctx	51.4
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.2
AD 5 Sup Temporal Ctx	34.9	Control (Path) 3 Occipital Ctx	2.4
AD 6 Inf Temporal Ctx	54.7	Control (Path) 4 Occipital Ctx	15.9
AD 6 Sup Temporal Ctx	60.3	Control 1 Parietal Ctx	8.4
Control 1 Temporal Ctx	3.7	Control 2 Parietal Ctx	36.3
Control 2 Temporal Ctx	20.3	Control 3 Parietal Ctx	15.9
Control 3 Temporal	9.7	Control (Path) 1	39.0

Ctx		Parietal Ctx	
Control 3 Temporal Ctx	7.2	Control (Path) 2 Parietal Ctx	16.2
Control (Path) 1 Temporal Ctx	46.7	Control (Path) 3 Parietal Ctx	2.2
Control (Path) 2 Temporal Ctx	34.9	Control (Path) 4 Parietal Ctx	36.1

Table BMC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3050, Run 167985384	Tissue Name	Rel. Exp.(%) Ag3050, Run 167985384
Liver adenocarcinoma	10.1	Kidney (fetal)	28.7
Pancreas	5.7	Renal ca. 786-0	14.9
Pancreatic ca. CAPAN 2	2.5	Renal ca. A498	6.6
Adrenal gland	2.5	Renal ca. RXF 393	8.8
Thyroid	3.7	Renal ca. ACHN	5.4
Salivary gland	1.5	Renal ca. UO-31	6.2
Pituitary gland	6.8	Renal ca. TK-10	6.5
Brain (fetal)	100.0	Liver	2.2
Brain (whole)	23.3	Liver (fetal)	6.4
Brain (amygdala)	22.2	Liver ca. (hepatoblast) HepG2	11.3
Brain (cerebellum)	88.9	Lung	0.8
Brain (hippocampus)	19.5	Lung (fetal)	15.1
Brain (substantia nigra)	19.3	Lung ca. (small cell) LX-1	14.4
Brain (thalamus)	7.7	Lung ca. (small cell) NCI-H69	18.4
Cerebral Cortex	12.1	Lung ca. (s.cell var.) SHP-77	43.2
Spinal cord	11.1	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	4.6	Lung ca. (non-sm. cell) A549	5.2
glio/astro U-118-MG	18.9	Lung ca. (non-s.cell) NCI-H23	7.4
astrocytoma SW1783	9.4	Lung ca. (non-s.cell) HOP-62	7.2
neuro*; met SK-N-AS	15.7	Lung ca. (non-s.cl) NCI-H522	21.0
astrocytoma SF-539	17.1	Lung ca. (squam.) SW 900	7.7
astrocytoma SNB-75	12.2	Lung ca. (squam.)	60.7

		NCI-H596	
glioma SNB-19	11.9	Mammary gland	1.3
glioma U251	15.7	Breast ca.* (pl.ef) MCF-7	14.5
glioma SF-295	16.4	Breast ca.* (pl.ef) MDA-MB-231	24.7
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	21.8
Heart	3.8	Breast ca. BT-549	7.3
Skeletal muscle (fetal)	8.1	Breast ca. MDA-N	13.5
Skeletal muscle	11.0	Ovary	4.0
Bone marrow	4.5	Ovarian ca. OVCAR-3	2.2
Thymus	9.6	Ovarian ca. OVCAR-4	4.7
Spleen	5.4	Ovarian ca. OVCAR-5	11.7
Lymph node	14.6	Ovarian ca. OVCAR-8	4.9
Colorectal	5.6	Ovarian ca. IGROV-1	2.2
Stomach	2.0	Ovarian ca.* (ascites) SK-OV-3	24.3
Small intestine	1.6	Uterus	6.9
Colon ca. SW480	5.1	Placenta	1.3
Colon ca.* SW620(SW480 met)	22.2	Prostate	5.9
Colon ca. HT29	4.9	Prostate ca.* (bone met)PC-3	4.8
Colon ca. HCT-116	9.3	Testis	11.2
Colon ca. CaCo-2	18.2	Melanoma Hs688(A).T	1.7
Colon ca. tissue(ODO3866)	2.4	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	10.1	Melanoma UACC-62	4.0
Gastric ca.* (liver met) NCI-N87	9.8	Melanoma M14	2.1
Bladder	6.3	Melanoma LOX IMVI	5.8
Trachea	2.2	Melanoma* (met) SK-MEL-5	2.1
Kidney	6.7	Adipose	4.3

PMA/ionomycin			
NK Cells IL-2 rest	14.3	NCI-H292 IL-13	5.2
Two Way MLR 3 day	7.5	NCI-H292 IFN gamma	9.3
Two Way MLR 5 day	7.0	HPAEC none	9.7
Two Way MLR 7 day	6.0	HPAEC TNF alpha + IL-1 beta	13.9
PBMC rest	3.4	Lung fibroblast none	5.3
PBMC PWM	38.4	Lung fibroblast TNF alpha + IL-1 beta	5.7
PBMC PHA-L	14.5	Lung fibroblast IL-4	7.7
Ramos (B cell) none	24.8	Lung fibroblast IL-9	7.0
Ramos (B cell) ionomycin	48.0	Lung fibroblast IL-13	3.8
B lymphocytes PWM	54.7	Lung fibroblast IFN gamma	10.4
B lymphocytes CD40L and IL-4	43.2	Dermal fibroblast CCD1070 rest	12.6
EOL-1 dbcAMP	10.7	Dermal fibroblast CCD1070 TNF alpha	30.1
EOL-1 dbcAMP PMA/ionomycin	10.8	Dermal fibroblast CCD1070 IL-1 beta	2.1
Dendritic cells none	6.9	Dermal fibroblast IFN gamma	4.5
Dendritic cells LPS	5.6	Dermal fibroblast IL-4	15.8
Dendritic cells anti-CD40	5.5	IBD Colitis 2	0.9
Monocytes rest	9.9	IBD Crohn's	1.2
Monocytes LPS	5.0	Colon	3.2
Macrophages rest	10.4	Lung	2.6
Macrophages LPS	2.0	Thymus	8.6
HUVEC none	8.1	Kidney	100.0
HUVEC starved	23.7		

Table BME. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag3050, Run 171694540	Tissue Name	Rel. Exp.(%) Ag3050, Run 171694540
BA4 Control	20.7	BA17 PSP	30.4
BA4 Control2	35.8	BA17 PSP2	15.7
BA4 Alzheimer's2	0.0	Sub Nigra Control	38.2
BA4 Parkinson's	45.7	Sub Nigra Control2	14.3
BA4 Parkinson's2	59.9	Sub Nigra Alzheimer's2	10.4

BA4 Huntington's	24.7	Sub Nigra Parkinson's2	46.7
BA4 Huntington's2	4.8	Sub Nigra Huntington's	49.0
BA4 PSP	31.4	Sub Nigra Huntington's2	11.6
BA4 PSP2	32.3	Sub Nigra PSP2	6.7
BA4 Depression	1.3	Sub Nigra Depression	9.2
BA4 Depression2	17.3	Sub Nigra Depression2	15.2
BA7 Control	38.2	Glob Palladus Control	23.2
BA7 Control2	1.2	Glob Palladus Control2	7.5
BA7 Alzheimer's2	5.4	Glob Palladus Alzheimer's	6.1
BA7 Parkinson's	43.5	Glob Palladus Alzheimer's2	0.0
BA7 Parkinson's2	57.8	Glob Palladus Parkinson's	100.0
BA7 Huntington's	43.8	Glob Palladus Parkinson's2	20.2
BA7 Huntington's2	59.5	Glob Palladus PSP	0.0
BA7 PSP	42.6	Glob Palladus PSP2	3.0
BA7 PSP2	31.9	Glob Palladus Depression	16.4
BA7 Depression	12.6	Temp Pole Control	18.9
BA9 Control	26.2	Temp Pole Control2	45.7
BA9 Control2	19.8	Temp Pole Alzheimer's	2.6
BA9 Alzheimer's	6.6	Temp Pole Alzheimer's2	10.0
BA9 Alzheimer's2	13.5	Temp Pole Parkinson's	22.8
BA9 Parkinson's	32.1	Temp Pole Parkinson's2	16.7
BA9 Parkinson's2	30.1	Temp Pole Huntington's	34.9
BA9 Huntington's	38.2	Temp Pole PSP	5.8
BA9 Huntington's2	31.6	Temp Pole PSP2	10.7
BA9 PSP	17.6	Temp Pole Depression2	24.3

BA9 PSP2	2.8	Cing Gyr Control	62.9
BA9 Depression	20.3	Cing Gyr Control2	13.4
BA9 Depression2	11.9	Cing Gyr Alzheimer's	23.7
BA17 Control	65.5	Cing Gyr Alzheimer's2	11.7
BA17 Control2	31.4	Cing Gyr Parkinson's	36.6
BA17 Alzheimer's2	11.0	Cing Gyr Parkinson's2	67.4
BA17 Parkinson's	62.9	Cing Gyr Huntington's	89.5
BA17 Parkinson's2	59.9	Cing Gyr Huntington's2	26.4
BA17 Huntington's	72.2	Cing Gyr PSP	11.6
BA17 Huntington's2	38.7	Cing Gyr PSP2	2.0
BA17 Depression	12.2	Cing Gyr Depression	13.0
BA17 Depression2	42.9	Cing Gyr Depression2	20.0

CNS_neurodegeneration_v1.0 Summary: Ag3050 This panel does not show differential expression of the NOV72 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3050 The NOV72 gene exhibits brain-preferential expression and is a member of a family of proteins that mediates ubiquitin-mediated protein degradation. Misprocessing of proteins involved in ubiquitin-mediated protein degradation is thought to be the cause of many neurodegenerative disorders such as Parkinson's disease, as well as those resulting from CAG repeat expansion genes, such as Huntington's disease. Therefore, therapeutic modulation of the expression or function of this gene may affect the protein degradation dysfunction seen in these diseases.

In addition, this gene is expressed at a slightly higher level in cancer cell lines compared to the normal lung, ovary, breast, and colon samples on this panel. This suggests that expression of this gene could be used as a diagnostic marker of cancer. Furthermore, inhibition of this gene product using small molecule drugs may be useful in the treatment of cancer in these tissues.

Among tissues with metabolic function, this gene is has a low level of expression in pancreas, thyroid, pituitary, heart, skeletal muscle, and adipose. This gene product may be a

small molecule target for the treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

Panel 4D Summary: Ag3050 The NOV72 gene is expressed at moderate to low levels (CT=29-34) in a wide range of cell types and tissues of significance in the immune response in health and disease, Highest expression of this gene is seen in kidney tissue (CT=29.36). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

Panel CNS_1 Summary: Ag3050 This panel confirms expression of the NOV72 gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

NOV73

Expression of gene NOV73 was assessed using the primer-probe set Ag3030, described in Table BNA.

Table BNA. Probe Name Ag3030

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5' - tgaaattcagaaccaggaaatg - 3'	22	838	1285
Probe	TET-5' - aaagagtgccttagcaggcacctccct - 3' - TAMRA	26	872	1286
Reverse	5' - aacctggcaatatgattcata - 3'	22	914	1287

CNS_neurodegeneration_v1.0 Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

Panel 2D Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 3D Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 5 Islet Summary: Ag3030 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

NOV74

Expression of gene NOV74 was assessed using the primer-probe set Ag3016, described in Table BOA. Results of the RTQ-PCR runs are shown in Tables BOB, BOC, BOD and BOE.

Table BOA. Probe Name Ag3016

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atctcagtgacctgctctcaga-3'	22	83	1288
Probe	TET-5'-cagtggctgctacagcctcccaag-3'-TAMRA	24	108	1289
Reverse	5'-aaatcttcagggtgacctcatt-3'	22	142	1290

Table BOB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3016, Run 209820675	Tissue Name	Rel. Exp.(%) Ag3016, Run 209820675
AD 1 Hippo	4.6	Control (Path) 3 Temporal Ctx	3.1
AD 2 Hippo	7.9	Control (Path) 4 Temporal Ctx	33.9
AD 3 Hippo	2.9	AD 1 Occipital Ctx	17.9
AD 4 Hippo	3.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	7.5
AD 6 Hippo	8.5	AD 4 Occipital Ctx	12.5
Control 2 Hippo	5.4	AD 5 Occipital Ctx	11.7
Control 4 Hippo	1.8	AD 6 Occipital Ctx	12.9
Control (Path) 3 Hippo	1.5	Control 1 Occipital Ctx	3.7
AD 1 Temporal Ctx	10.8	Control 2 Occipital Ctx	18.8
AD 2 Temporal Ctx	20.4	Control 3 Occipital Ctx	0.7
AD 3 Temporal Ctx	3.5	Control 4 Occipital Ctx	5.8
AD 4 Temporal Ctx	22.2	Control (Path) 1 Occipital Ctx	50.3
AD 5 Inf Temporal Ctx	60.7	Control (Path) 2 Occipital Ctx	23.3
AD 5 Sup Temporal	17.7	Control (Path) 3	1.5

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	23.7	Control (Path) 4 Occipital Ctx	36.6
AD 6 Sup Temporal Ctx	23.7	Control 1 Parietal Ctx	6.1
Control 1 Temporal Ctx	5.6	Control 2 Parietal Ctx	28.7
Control 2 Temporal Ctx	5.9	Control 3 Parietal Ctx	21.3
Control 3 Temporal Ctx	11.1	Control (Path) 1 Parietal Ctx	49.3
Control 3 Temporal Ctx	6.3	Control (Path) 2 Parietal Ctx	27.0
Control (Path) 1 Temporal Ctx	45.4	Control (Path) 3 Parietal Ctx	5.6
Control (Path) 2 Temporal Ctx	36.9	Control (Path) 4 Parietal Ctx	39.0

Table BOC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3016, Run 167819111	Tissue Name	Rel. Exp.(%) Ag3016, Run 167819111
Liver adenocarcinoma	0.0	Kidney (fetal)	9.2
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	99.3	Liver	0.0
Brain (whole)	96.6	Liver (fetal)	0.0
Brain (amygdala)	21.5	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	12.0	Lung	0.0
Brain (hippocampus)	24.7	Lung (fetal)	6.2
Brain (substantia nigra)	15.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	8.2
Cerebral Cortex	100.0	Lung ca. (s.cell var.) SHP-77	6.9
Spinal cord	0.0	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	46.0	Lung ca. (non-sm.	10.9

		cell) A549	
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	6.0
astrocytoma SW1783	6.3	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	7.1	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.0	Mammary gland	0.0
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	32.5	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl.ef) T47D	81.2
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	10.6	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	7.0	Ovarian ca. OVCAR-4	4.5
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	6.3	Ovarian ca. OVCAR-8	100.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	72.7
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	15.4
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	19.5
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	57.0

Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	26.8
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	19.8
Kidney	0.0	Adipose	0.0

Table BOD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3016, Run 164404251	Tissue Name	Rel. Exp.(%) Ag3016, Run 164404251
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	6.2	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	13.7
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-	0.0	CCD1106 (Keratinocytes)	0.0

CD95 CH11		none	
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	56.6
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	0.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	7.6
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	6.0
Dendritic cells anti- CD40	0.0	IBD Colitis 2	5.6
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	0.0	Lung	6.7
Macrophages LPS	0.0	Thymus	4.3
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

Table BOE. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag3016, Run 171688428	Tissue Name	Rel. Exp.(%) Ag3016, Run 171688428
BA4 Control	9.9	BA17 PSP	28.5
BA4 Control2	12.0	BA17 PSP2	17.8
BA4 Alzheimer's2	2.9	Sub Nigra Control	5.4
BA4 Parkinson's	28.9	Sub Nigra Control2	0.0
BA4 Parkinson's2	35.1	Sub Nigra Alzheimer's2	7.7
BA4 Huntington's	12.2	Sub Nigra Parkinson's2	15.1
BA4 Huntington's2	8.3	Sub Nigra Huntington's	3.5
BA4 PSP	11.4	Sub Nigra Huntington's2	0.0
BA4 PSP2	17.9	Sub Nigra PSP2	0.0
BA4 Depression	9.7	Sub Nigra Depression	5.2
BA4 Depression2	11.7	Sub Nigra Depression2	14.8
BA7 Control	31.4	Glob Palladus Control	5.0
BA7 Control2	11.7	Glob Palladus Control2	0.0
BA7 Alzheimer's2	20.2	Glob Palladus Alzheimer's	0.0
BA7 Parkinson's	15.3	Glob Palladus Alzheimer's2	10.6
BA7 Parkinson's2	59.0	Glob Palladus Parkinson's	58.2
BA7 Huntington's	43.2	Glob Palladus Parkinson's2	0.0
BA7 Huntington's2	46.0	Glob Palladus PSP	0.0
BA7 PSP	37.6	Glob Palladus PSP2	5.0
BA7 PSP2	21.9	Glob Palladus Depression	5.7
BA7 Depression	21.9	Temp Pole Control	5.9
BA9 Control	14.8	Temp Pole Control2	0.0
BA9 Control2	24.5	Temp Pole Alzheimer's	0.0
BA9 Alzheimer's	10.2	Temp Pole Alzheimer's2	9.4
BA9	12.2	Temp Pole	30.4

Alzheimer's2		Parkinson's	
BA9 Parkinson's	52.1	Temp Pole Parkinson's2	37.1
BA9 Parkinson's2	45.4	Temp Pole Huntington's	11.5
BA9 Huntington's	20.2	Temp Pole PSP	19.3
BA9 Huntington's2	35.1	Temp Pole PSP2	0.0
BA9 PSP	24.3	Temp Pole Depression2	12.5
BA9 PSP2	9.5	Cing Gyr Control	21.0
BA9 Depression	9.8	Cing Gyr Control2	16.2
BA9 Depression2	12.2	Cing Gyr Alzheimer's	0.0
BA17 Control	100.0	Cing Gyr Alzheimer's2	0.0
BA17 Control2	40.9	Cing Gyr Parkinson's	10.5
BA17 Alzheimer's2	28.9	Cing Gyr Parkinson's2	0.0
BA17 Parkinson's	94.6	Cing Gyr Huntington's	35.4
BA17 Parkinson's2	28.1	Cing Gyr Huntington's2	5.0
BA17 Huntington's	25.9	Cing Gyr PSP	0.0
BA17 Huntington's2	19.3	Cing Gyr PSP2	9.8
BA17 Depression	33.4	Cing Gyr Depression	5.0
BA17 Depression2	56.3	Cing Gyr Depression2	24.8

CNS_neurodegeneration_v1.0 Summary: Ag3016 This panel does not show differential expression of the NOV74 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3016 The NOV74 gene represents a dual specificity phosphatase that is expressed preferentially at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently,

agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

In addition, there are low but significant levels of expression in samples derived from breast cancer, ovarian cancer, and melanoma cell lines. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of breast cancer, ovarian cancer, and melanoma.

References:

Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. *Brain Res Mol Brain Res* 1995 Feb;28(2):353-6

PAC-1 mRNA has previously been found only in activated T-cells in vitro and in vivo. The gene encodes a dual specificity protein phosphatase that regulates MAP kinase activity. Here, I describe that PAC-1 mRNA is induced also in neurons in the rat brain following 30 min of forebrain ischemia. At 6, 12 and 24 h after ischemia, PAC-1 mRNA was found most prominently in hippocampal cells which are resistant to 30 min of forebrain ischemia, but not in the selectively vulnerable CA1 sector. At later time points and in control animals no PAC-1 mRNA could be detected in any brain region. The protein-tyrosine/threonine phosphatase PAC-1, therefore, may be involved in adaptational responses of hippocampal cells resistant to ischemic injury.

Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. *Neuroreport* 1997 Sep 29;8(14):3077-80

Recurrent seizure activity leads to delayed neuronal death as well as to inflammatory responses involving microglia in hippocampal subfields CA1, CA3 and CA4. Since mitogen activated protein (MAP) kinases control neuronal apoptosis and trigger generation of inflammatory cytokines, their activation state could determine seizure-related brain damage. PAC1 is a dual specificity protein phosphatase inactivating MAP kinases which we have found to be undetectable in normal brain. Despite this, kainic acid-induced seizure activity lead to rapid (approximately 3 h) but transient appearance of PAC1 mRNA in granule cells of the dentate gyrus as well as in pyramidal CA1 neurons. This pattern changed with time and after 2-3 days PAC1 was induced in dying CA1 and CA3 neurons. At this time PAC1 mRNA was also expressed in white matter microglia as well as in microglia invading the damaged

hippocampus. PAC1 may play an important role controlling MAP kinase involvement in both neuronal death and neuro-inflammation following excitotoxic damage.

Panel 4D Summary: Ag3016 The NOV74 gene is only expressed at detectable levels in the kidney (CT = 34.2) among the samples on this panel. Thus, expression of this gene could be used to distinguish kidney from the other samples on this panel. In addition, the dual-specificity protein phosphatase encoded for by this gene could allow cells within the kidney to respond to specific microenvironmental signals. Furthermore, small molecule therapies designed with the protein encoded for by this gene could modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

Panel CNS_1 Summary: Ag3016 This panel confirms expression of the NOV74 gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

NOV75

Expression of gene NOV75 was assessed using the primer-probe sets Ag3020 and Ag2968, described in Tables BPA and BPB. Results of the RTQ-PCR runs are shown in Tables BPC, BPD, BPE, BPF and BPG.

Table BPA. Probe Name Ag3020

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaatcaccacattctgaat-3'	21	265	1291
Probe	TET-5'-cgtttacactggccccgaattctaca-3'-TAMRA	26	303	1292
Reverse	5'-cctctacaccaggtactggat-3'	22	340	1293

Table BPB. Probe Name Ag2968

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggaatcaccacattctgaat-3'	21	265	1294
Probe	TET-5'-cgtttacactggccccgaattctaca-3'-TAMRA	26	303	1295
Reverse	5'-cctctacaccaggtactggat-3'	22	340	1296

Table BPC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3020, Run 221998694	Tissue Name	Rel. Exp.(%) Ag3020, Run 221998694
Adipose	0.4	Renal ca. TK-10	0.0

Melanoma* Hs688(A).T	0.0	Bladder	0.1
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK- MEL-5	0.0	Colon ca. SW480	0.2
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.2	Colon ca. CaCo-2	0.0
Placenta	0.1	Colon cancer tissue	31.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	100.0
Ovarian ca. SK-OV- 3	0.3	Colon ca. SW-48	5.7
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.1	Small Intestine Pool	0.0
Ovarian ca. IGROV- 1	18.9	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.1	Fetal Heart	12.7
Breast ca. MCF-7	0.0	Heart Pool	4.9
Breast ca. MDA- MB-231	0.0	Lymph Node Pool	0.1
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	23.3
Breast ca. T47D	0.0	Skeletal Muscle Pool	20.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.3	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.1
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF- 539	0.7

Lung ca. LX-1	12.7	CNS cancer (astro) SNB-75	0.1
Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	15.9
Lung ca. SHP-77	0.0	CNS cancer (glio) SF- 295	0.5
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.1
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.1
Lung ca. HOP-62	0.1	Cerebral Cortex Pool	0.1
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.1
Liver	0.0	Brain (Thalamus) Pool	0.1
Fetal Liver	0.1	Brain (whole)	0.1
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.1	Adrenal Gland	0.0
Fetal Kidney	0.1	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.4
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table BPD. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2968, Run 166220058	Rel. Exp.(%) Ag3020, Run 167819114	Tissue Name	Rel. Exp.(%) Ag2968, Run 166220058	Rel. Exp.(%) Ag3020, Run 167819114
Liver adenocarcinoma	0.0	0.1	Kidney (fetal)	0.0	0.3
Pancreas	0.0	0.0	Renal ca. 786- 0	0.0	0.0
Pancreatic ca. CAPAN 2	0.1	0.2	Renal ca. A498	0.0	0.0
Adrenal gland	0.0	0.0	Renal ca. RXF 393	0.0	0.0
Thyroid	0.3	1.6	Renal ca. ACHN	0.0	0.0
Salivary gland	0.8	0.6	Renal ca. UO- 31	0.0	0.0
Pituitary gland	0.0	0.0	Renal ca. TK- 10	0.0	0.0
Brain (fetal)	0.0	0.1	Liver	0.0	0.0
Brain (whole)	0.4	1.0	Liver (fetal)	0.5	0.0

Brain (amygdala)	0.2	1.0	Liver ca. (hepatoblast) HepG2	0.0	0.0
Brain (cerebellum)	0.0	0.0	Lung	0.0	0.5
Brain (hippocampus)	0.0	0.0	Lung (fetal)	0.0	0.0
Brain (substantia nigra)	0.2	0.1	Lung ca. (small cell) LX-1	10.7	16.2
Brain (thalamus)	0.0	0.0	Lung ca. (small cell) NCI-H69	0.0	0.5
Cerebral Cortex	0.4	0.1	Lung ca. (s.cell var.) SHP-77	0.0	0.0
Spinal cord	0.0	0.0	Lung ca. (large cell)NCI-H460	0.0	0.0
glio/astro U87-MG	0.0	0.0	Lung ca. (non- sm. cell) A549	0.0	0.2
glio/astro U-118- MG	0.0	0.1	Lung ca. (non- s.cell) NCI- H23	0.0	0.0
astrocytoma SW1783	0.0	0.0	Lung ca. (non- s.cell) HOP-62	0.1	0.1
neuro*; met SK-N- AS	0.0	0.0	Lung ca. (non- s.cl) NCI- H522	0.0	0.0
astrocytoma SF- 539	12.1	8.4	Lung ca. (squam.) SW 900	0.0	0.0
astrocytoma SNB- 75	0.0	0.3	Lung ca. (squam.) NCI- H596	0.1	0.4
glioma SNB-19	0.0	0.0	Mammary gland	0.2	0.2
glioma U251	0.4	0.0	Breast ca.* (pl.ef) MCF-7	0.0	0.0
glioma SF-295	0.3	0.3	Breast ca.* (pl.ef) MDA- MB-231	0.0	0.0
Heart (fetal)	7.4	26.8	Breast ca.* (pl.ef) T47D	0.0	0.0
Heart	29.9	35.4	Breast ca. BT- 549	0.0	0.0
Skeletal muscle (fetal)	10.8	33.9	Breast ca. MDA-N	0.0	0.0
Skeletal muscle	100.0	100.0	Ovary	0.0	0.1

Bone marrow	0.1	0.6	Ovarian ca. OVCAR-3	0.0	0.0
Thymus	0.1	0.1	Ovarian ca. OVCAR-4	0.0	0.0
Spleen	0.0	0.0	Ovarian ca. OVCAR-5	0.0	0.6
Lymph node	0.0	0.0	Ovarian ca. OVCAR-8	0.0	0.0
Colorectal	0.2	0.2	Ovarian ca. IGROV-1	26.2	26.2
Stomach	0.0	0.0	Ovarian ca.* (ascites) SK-OV-3	0.2	1.0
Small intestine	0.0	0.0	Uterus	0.0	0.0
Colon ca. SW480	0.0	0.2	Placenta	1.0	0.0
Colon ca.* SW620(SW480 met)	1.6	6.1	Prostate	0.2	0.1
Colon ca. HT29	0.0	0.0	Prostate ca.* (bone met)PC-3	0.0	0.0
Colon ca. HCT-116	0.0	0.0	Testis	0.2	0.2
Colon ca. CaCo-2	0.0	0.0	Melanoma Hs688(A).T	0.0	0.0
Colon ca. tissue(ODO3866)	21.9	30.6	Melanoma* (met) Hs688(B).T	0.0	0.0
Colon ca. HCC-2998	0.0	0.5	Melanoma UACC-62	0.0	0.0
Gastric ca.* (liver met) NCI-N87	0.0	0.0	Melanoma M14	0.0	0.0
Bladder	0.1	0.0	Melanoma LOX IMVI	0.0	0.0
Trachea	0.8	0.6	Melanoma* (met) SK-MEL-5	0.0	0.0
Kidney	0.0	0.0	Adipose	1.1	1.7

Table BPE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag2968, Run 170188142	Tissue Name	Rel. Exp.(%) Ag2968, Run 170188142
Daoy- Medulloblastoma	0.0	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0

TE671- Medulloblastoma	2.2	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	0.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	0.0	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	0.0
SNB-78- Glioma	0.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	0.0	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	0.0	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	0.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	0.0	JM1- pre-B-cell lymphoma	0.0
Cerebellum	0.0	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	0.0	TF-1- Erythroleukemia	0.0
DMS-114- Small cell lung cancer	0.0	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	0.0	U937- Histiocytic lymphoma	0.0
NCI-H146- Small cell lung cancer	0.0	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	0.0
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	0.0	G401- Wilms' tumor	0.0
NCI-H1155- Large cell lung cancer	0.0	Hs766T- Pancreatic carcinoma (LN metastasis)	0.0
NCI-H1299- Large cell lung cancer	0.0	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.0
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.0

LX-1- Small cell lung cancer	4.4	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	100.0	MIA PaCa-2- Pancreatic carcinoma	0.0
KM12- Colon cancer	0.0	CFPAC-1- Pancreatic ductal adenocarcinoma	0.0
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	0.0
NCI-H716- Colon cancer	0.0	T24- Bladder carcinma (transitional cell)	0.0
SW-48- Colon adenocarcinoma	4.3	5637- Bladder carcinoma	0.0
SW1116- Colon adenocarcinoma	0.0	HT-1197- Bladder carcinoma	0.0
LS 174T- Colon adenocarcinoma	10.1	UM-UC-3- Bladder carcinma (transitional cell)	0.0
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	0.0
SW-480- Colon adenocarcinoma	0.0	HT-1080- Fibrosarcoma	0.0
NCI-SNU-5- Gastric carcinoma	0.0	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	1.2	SK-LMS-1- Leiomyosarcoma (vulva)	0.0
NCI-SNU-16- Gastric carcinoma	0.0	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.7
NCI-SNU-1- Gastric carcinoma	0.1	A431- Epidermoid carcinoma	0.0
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	0.0
RF-48- Gastric adenocarcinoma	0.0	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	0.0	MDA-MB-468- Breast adenocarcinoma	0.0
NCI-N87- Gastric carcinoma	0.0	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	0.0	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.0
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.0

Table BPF. Panel 4D

Tissue Name	Rel. Exp.(%)	Tissue Name	Rel. Exp.(%)
-------------	--------------	-------------	--------------

	Ag3020, Run 164528102		Ag3020, Run 164528102
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	65.1
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	21.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	10.7
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1 anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	13.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	100.0
LAK cells IL-2+IL-12	0.0	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	8.0	NCI-H292 IL-13	0.0

Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	10.6	HPAEC TNF alpha + IL-1 beta	11.2
PBMC rest	0.0	Lung fibroblast none	0.0
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.0
B lymphocytes PWM	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	9.9	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	23.8	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	0.0
Monocytes LPS	0.0	Colon	44.4
Macrophages rest	0.0	Lung	26.6
Macrophages LPS	0.0	Thymus	0.0
HUVEC none	0.0	Kidney	0.0
HUVEC starved	0.0		

Table BPG. Panel 5D

Tissue Name	Rel. Exp.(%) Ag2968, Run 169270971	Rel. Exp.(%) Ag3020, Run 172171108	Tissue Name	Rel. Exp.(%) Ag2968, Run 169270971	Rel. Exp.(%) Ag3020, Run 172171108
97457_Patient-02go_adipose	0.0	0.2	94709_Donor 2 AM - A_adipose	0.0	0.0
97476_Patient-07sk_skeletal muscle	1.2	8.4	94710_Donor 2 AM - B_adipose	0.0	0.0
97477_Patient-07ut_uterus	0.1	0.0	94711_Donor 2 AM - C_adipose	0.0	0.0

97478_Patient-07pl_placenta	0.0	1.3	94712_Donor 2 AD - A_adipose	0.0	3.0
97481_Patient-08sk_skeletal muscle	2.8	12.6	94713_Donor 2 AD - B_adipose	100.0	0.0
97482_Patient-08ut_uterus	0.0	0.0	94714_Donor 2 AD - C_adipose	0.0	0.0
97483_Patient-08pl_placenta	0.0	0.0	94742_Donor 3 U - A Mesenchymal Stem Cells	0.0	0.0
97486_Patient-09sk_skeletal muscle	3.0	12.9	94743_Donor 3 U - B_Mesenchymal Stem Cells	0.0	0.0
97487_Patient-09ut_uterus	0.1	0.7	94730_Donor 3 AM - A_adipose	0.0	0.0
97488_Patient-09pl_placenta	0.1	0.2	94731_Donor 3 AM - B_adipose	0.0	0.0
97492_Patient-10ut_uterus	0.0	0.2	94732_Donor 3 AM - C_adipose	0.0	0.0
97493_Patient-10pl_placenta	0.2	1.1	94733_Donor 3 AD - A_adipose	0.0	0.0
97495_Patient-11go_adipose	0.0	0.0	94734_Donor 3 AD - B_adipose	0.0	0.0
97496_Patient-11sk_skeletal muscle	8.5	53.6	94735_Donor 3 AD - C_adipose	0.0	0.0
97497_Patient-11ut_uterus	0.0	0.3	77138_Liver_HepG2untreated	0.0	0.0
97498_Patient-11pl_placenta	0.1	2.3	73556_Heart_Cardiac stromal cells (primary)	0.0	0.0
97500_Patient-12go_adipose	0.0	0.4	81735_Small Intestine	0.1	0.0
97501_Patient-12sk_skeletal muscle	19.1	100.0	72409_Kidney_Proximal Convoluted Tubule	0.0	0.0
97502_Patient-12ut_uterus	0.1	0.0	82685_Small intestine_Duodenum	0.0	0.0
97503_Patient-12pl_placenta	0.1	1.3	90650_Adrenal_Adrenocortical adenoma	0.0	0.2
94721_Donor 2 U - A_Mesenchymal Stem Cells	0.0	0.0	72410_Kidney_HRCE	0.0	0.3
94722_Donor 2 U - B_Mesenchymal Stem Cells	0.0	0.0	72411_Kidney_HRE	0.0	0.0
94723_Donor 2	0.0	0.0	73139_Uterus_Uterine smooth	0.0	0.0

U - C_Mesenchymal Stem Cells			muscle cells		
------------------------------------	--	--	--------------	--	--

General_screening_panel_v1.4 Summary: Ag3020 The NOV75 gene is expressed in brain, colon, lung and ovarian cancer cell lines with highest expression in a colon cancer cell line Colo-205 (CT=24.37). This suggests that this gene can be used as a diagnostic marker for these types of cancer. Furthermore, inhibition of the protein using small molecule drugs could potentially be useful for the treatment of brain, colon, lung and ovarian cancer.

In addition, this gene has low expression in adipose and high expression in adult and fetal heart and skeletal muscle. Thus, this protein phosphatase may be a small molecule target for the treatment of obesity, Type 2 diabetes and cardiac and skeletal muscle disease.

Panel 1.3D Summary: Ag2968/Ag3020 Results from two experiments using identical probe/primer sets are in excellent agreement. Expression of the NOV75 gene is highest in adult skeletal muscle (CTs = 26-28). Significant but somewhat lower expression is also seen in fetal skeletal muscle and adult/fetal heart. Thus, expression of this gene may be used to distinguish these samples from the other samples on this panel.

This gene is also expressed in brain, colon, lung and ovarian cancer cell lines, consistent with General_screening_panel_v1.4. This suggests that this gene can be used as a diagnostic marker for these types of cancer and inhibition of the protein using small molecule drugs can be used for the treatment of brain, colon, lung and ovarian cancer.

Panel 3D Summary: Ag2968 Expression of the NOV75 gene is highest in colon cancer cell line Colo-205 (CT = 25.6). In addition, significant expression of this gene is seen in two other colon cancer cell lines. Thus, expression of this gene may be used to distinguish these colon cancer cell lines from the other samples on this panel. Moreover, therapeutic modulation of the activity of this gene or its protein product, using small molecules, antibodies or protein therapeutics, may be of benefit in the treatment of colon cancer.

Panel 4D Summary: Ag3020 Expression of the NOV75 gene is highest in a liver cirrhosis sample (CT = 33.3). Furthermore, expression of this gene is not detected in normal liver in Panels 1.3D or 1.4, suggesting that its expression is unique to liver cirrhosis. This gene encodes a putative protein phosphatase; therefore, antibodies or small molecule therapeutics could reduce or inhibit fibrosis that occurs in liver cirrhosis. In addition, antibodies to this protein could also be used for the diagnosis of liver cirrhosis. Low levels of expression are also seen in colon and resting astrocytes.

$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ -1 & i \end{pmatrix}$, $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$

5

NOV76a

10

Table BQA. Probe Name Ag3022

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgcttgcaggtttattcttagg-3'	22	431	1297
Probe	TET-5'-tgctgagttctctcgttggttccctg-3'-TAMRA	26	459	1298
Reverse	5'-tgagaaatgcaggtaggga-3'	22	509	1299

Table BQB. Probe Name Ag4891

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aatacctgtccaaagcctgact-3'	22	661	1300
Probe	TET-5'-ttatccccgagtgctcatttctgcgt-3'-TAMRA	26	683	1301
Reverse	5'-ttctcacaaaagctgtcattca-3'	22	716	1302

15

Tissue Name	Rel. Exp.(%) Ag4891, Run 228714709	Tissue Name	Rel. Exp.(%) Ag4891, Run 228714709
Adipose	10.8	Renal ca. TK-10	24.5
Melanoma* Hs688(A).T	5.6	Bladder	22.8
Melanoma* Hs688(B).T	6.3	Gastric ca. (liver met.) NCI-N87	46.7
Melanoma* M14	27.4	Gastric ca. KATO III	22.2
Melanoma* LOXIMVI	9.5	Colon ca. SW-948	11.5
Melanoma* SK- MEL-5	12.4	Colon ca. SW480	20.7

Squamous cell carcinoma SCC-4	10.2	Colon ca.* (SW480 met) SW620	11.6
Testis Pool	11.3	Colon ca. HT29	10.6
Prostate ca.* (bone met) PC-3	6.7	Colon ca. HCT-116	12.2
Prostate Pool	16.8	Colon ca. CaCo-2	73.2
Placenta	33.7	Colon cancer tissue	45.4
Uterus Pool	9.7	Colon ca. SW1116	3.0
Ovarian ca. OVCAR-3	30.4	Colon ca. Colo-205	5.4
Ovarian ca. SK-OV-3	10.2	Colon ca. SW-48	10.2
Ovarian ca. OVCAR-4	8.5	Colon Pool	15.9
Ovarian ca. OVCAR-5	21.8	Small Intestine Pool	13.2
Ovarian ca. IGROV-1	10.6	Stomach Pool	7.4
Ovarian ca. OVCAR-8	5.6	Bone Marrow Pool	5.7
Ovary	6.5	Fetal Heart	11.3
Breast ca. MCF-7	52.5	Heart Pool	7.0
Breast ca. MDA-MB-231	11.5	Lymph Node Pool	18.4
Breast ca. BT 549	41.8	Fetal Skeletal Muscle	9.2
Breast ca. T47D	4.4	Skeletal Muscle Pool	16.3
Breast ca. MDA-N	12.1	Spleen Pool	19.3
Breast Pool	17.8	Thymus Pool	14.7
Trachea	12.1	CNS cancer (glio/astro) U87-MG	12.2
Lung	4.7	CNS cancer (glio/astro) U-118-MG	11.8
Fetal Lung	36.3	CNS cancer (neuro;met) SK-N-AS	24.5
Lung ca. NCI-N417	4.4	CNS cancer (astro) SF-539	5.4
Lung ca. LX-1	9.9	CNS cancer (astro) SNB-75	13.0
Lung ca. NCI-H146	15.0	CNS cancer (glio) SNB-19	11.7
Lung ca. SHP-77	6.8	CNS cancer (glio) SF-295	15.7
Lung ca. A549	28.7	Brain (Amygdala) Pool	10.8
Lung ca. NCI-H526	0.8	Brain (cerebellum)	35.8
Lung ca. NCI-H23	38.4	Brain (fetal)	9.3

Lung ca. NCI-H460	10.0	Brain (Hippocampus) Pool	10.2
Lung ca. HOP-62	11.7	Cerebral Cortex Pool	9.7
Lung ca. NCI-H522	29.5	Brain (Substantia nigra) Pool	10.3
Liver	6.1	Brain (Thalamus) Pool	15.1
Fetal Liver	19.3	Brain (whole)	10.3
Liver ca. HepG2	34.4	Spinal Cord Pool	10.3
Kidney Pool	23.7	Adrenal Gland	100.0
Fetal Kidney	11.6	Pituitary gland Pool	5.3
Renal ca. 786-0	14.6	Salivary Gland	12.5
Renal ca. A498	5.4	Thyroid (female)	6.7
Renal ca. ACHN	21.6	Pancreatic ca. CAPAN2	21.0
Renal ca. UO-31	25.9	Pancreas Pool	25.9

Table BQD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4891, Run 214253687	Tissue Name	Rel. Exp.(%) Ag4891, Run 214253687
Secondary Th1 act	55.1	HUVEC IL-1beta	17.7
Secondary Th2 act	100.0	HUVEC IFN gamma	6.2
Secondary Tr1 act	74.2	HUVEC TNF alpha + IFN gamma	9.0
Secondary Th1 rest	18.8	HUVEC TNF alpha + IL4	11.7
Secondary Th2 rest	34.4	HUVEC IL-11	2.9
Secondary Tr1 rest	20.9	Lung Microvascular EC none	21.3
Primary Th1 act	18.8	Lung Microvascular EC TNFalpha + IL-1beta	26.2
Primary Th2 act	60.3	Microvascular Dermal EC none	6.7
Primary Tr1 act	35.4	Microvascular Dermal EC TNFalpha + IL-1beta	15.1
Primary Th1 rest	17.7	Bronchial epithelium TNFalpha + IL1beta	16.8
Primary Th2 rest	24.8	Small airway epithelium none	4.0
Primary Tr1 rest	14.8	Small airway epithelium TNFalpha + IL-1beta	13.3
CD45RA CD4 lymphocyte act	15.1	Coronary artery SMC rest	8.4
CD45RO CD4 lymphocyte act	63.3	Coronary artery SMC TNFalpha + IL-1beta	8.2
CD8 lymphocyte act	23.0	Astrocytes rest	8.4

Secondary CD8 lymphocyte rest	46.0	Astrocytes TNFalpha + IL-1beta	3.1
Secondary CD8 lymphocyte act	18.6	KU-812 (Basophil) rest	0.5
CD4 lymphocyte none	15.5	KU-812 (Basophil) PMA/ionomycin	2.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	47.0	CCD1106 (Keratinocytes) none	3.3
LAK cells rest	15.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	10.7
LAK cells IL-2	30.4	Liver cirrhosis	13.7
LAK cells IL-2+IL-12	12.9	NCI-H292 none	28.9
LAK cells IL-2+IFN gamma	12.8	NCI-H292 IL-4	24.0
LAK cells IL-2+ IL-18	22.4	NCI-H292 IL-9	52.9
LAK cells PMA/ionomycin	24.8	NCI-H292 IL-13	26.6
NK Cells IL-2 rest	20.0	NCI-H292 IFN gamma	30.8
Two Way MLR 3 day	29.7	HPAEC none	4.2
Two Way MLR 5 day	32.5	HPAEC TNF alpha + IL-1 beta	58.2
Two Way MLR 7 day	42.6	Lung fibroblast none	10.0
PBMC rest	10.3	Lung fibroblast TNF alpha + IL-1 beta	7.5
PBMC PWM	47.0	Lung fibroblast IL-4	4.5
PBMC PHA-L	49.7	Lung fibroblast IL-9	5.9
Ramos (B cell) none	1.6	Lung fibroblast IL-13	3.4
Ramos (B cell) ionomycin	1.5	Lung fibroblast IFN gamma	8.4
B lymphocytes PWM	28.1	Dermal fibroblast CCD1070 rest	4.1
B lymphocytes CD40L and IL-4	11.4	Dermal fibroblast CCD1070 TNF alpha	42.9
EOL-1 dbcAMP	1.5	Dermal fibroblast CCD1070 IL-1 beta	4.9
EOL-1 dbcAMP PMA/ionomycin	1.9	Dermal fibroblast IFN gamma	7.3
Dendritic cells none	10.2	Dermal fibroblast IL-4	6.8
Dendritic cells LPS	7.3	Dermal Fibroblasts rest	7.7
Dendritic cells anti-CD40	6.0	Neutrophils TNFa+LPS	13.5
Monocytes rest	2.5	Neutrophils rest	12.8
Monocytes LPS	34.9	Colon	6.0
Macrophages rest	9.4	Lung	10.6
Macrophages LPS	7.1	Thymus	20.3

HUVEC none	5.4	Kidney	10.2
HUVEC starved	8.5		

CNS_neurodegeneration_v1.0 Summary: Ag3022 No significant expression detected. The amp plot indicates that there is possibility of a potential chemistry or probe/primer failure (data not shown).

5 **General_screening_panel_v1.4 Summary:** Ag3022 The amp plot indicates that there is possibility of a potential chemistry or probe/primer failure (data not shown).

General_screening_panel_v1.5 Summary: Ag4891 The NOV76a gene has moderate levels of expression in adipose, liver, heart, skeletal muscle, pituitary, thyroid and pancreas, and high levels of expression in adrenal gland. Thus, this gene product may be a small
10 molecule target for the treatment of metabolic, endocrine and adrenal diseases, including obesity, Types 1 and 2 diabetes, and Addison's disease.

In addition, this gene is expressed at moderate levels in the cancer cell lines in this panel. A higher level of expression is observed in colon, lung, breast and ovarian cancer cell lines when compared to samples from the normal colon, lung, breast and ovary. Thus, this
15 gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be useful for the treatment of cancer in these tissues.

This gene encodes a homolog of a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of
20 MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

25 **References:**

Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. Brain Res Mol Brain Res 1995 Feb;28(2):353-6

PAC-1 mRNA has previously been found only in activated T-cells in vitro and in vivo.
30 The gene encodes a dual specificity protein phosphatase that regulates MAP kinase activity. Here, I describe that PAC-1 mRNA is induced also in neurons in the rat brain following 30 min of forebrain ischemia. At 6, 12 and 24 h after ischemia, PAC-1 mRNA was found most prominently in hippocampal cells which are resistant to 30 min of forebrain ischemia, but not

in the selectively vulnerable CA1 sector. At later time points and in control animals no PAC-1 mRNA could be detected in any brain region. The protein-tyrosine/threonine phosphatase PAC-1, therefore, may be involved in adaptational responses of hippocampal cells resistant to ischemic injury.

- 5 Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. Neuroreport 1997 Sep 29;8(14):3077-80

10 Recurrent seizure activity leads to delayed neuronal death as well as to inflammatory responses involving microglia in hippocampal subfields CA1, CA3 and CA4. Since mitogen activated protein (MAP) kinases control neuronal apoptosis and trigger generation of inflammatory cytokines, their activation state could determine seizure-related brain damage. PAC1 is a dual specificity protein phosphatase inactivating MAP kinases which we have found to be undetectable in normal brain. Despite this, kainic acid-induced seizure activity lead to rapid (approximately 3 h) but transient appearance of PAC1 mRNA in granule cells of the dentate gyrus as well as in pyramidal CA1 neurons. This pattern changed with time and after 2-3 days PAC1 was induced in dying CA1 and CA3 neurons. At this time PAC1 mRNA was also expressed in white matter microglia as well as in microglia invading the damaged hippocampus. PAC1 may play an important role controlling MAP kinase involvement in both neuronal death and neuro-inflammation following excitotoxic damage.

- 20 **Panel 4.1D Summary:** Ag4891 The NOV76a gene is expressed in a wide range of cell types and tissues (CT=26-34) of significance in the immune response in health and disease. Highest expression of this gene is detected in activated secondary Th2 cells (CT=26.48). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

25 **Panel 4D Summary:** Ag3022 No significant expression detected. Potential probe/primer failure (data not shown).

30 NOV77

 Expression of gene NOV77 was assessed using the primer-probe sets Ag3023 and Ag3373, described in Tables BRA and BRB. Results of the RTQ-PCR runs are shown in Tables BRC, BRD, BRE and BRF.

Table BRA. Probe Name Ag3023

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctaagtctggatttgcctca-3'	22	572	1303
Probe	TET-5'-tcaggaatatgaagccatctacctagca-3'-TAMRA	28	597	1304
Reverse	5'-tggagtgggtgacatcatctgta-3'	22	635	1305

Table BRB. Probe Name Ag3373

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atattgtccatcaacttcaggaa-3'	22	582	1306
Probe	TET-5'-tgaagccatctacctagcaaaattaaca-3'-TAMRA	28	606	1307
Reverse	5'-tggagtgggtgacatcatctgta-3'	22	635	1308

5

Table BRC. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071	Tissue Name	Rel. Exp.(%) Ag3023, Run 209821074	Rel. Exp.(%) Ag3373, Run 210154071
AD 1 Hippo	10.9	16.8	Control (Path) 3 Temporal Ctx	9.1	8.0
AD 2 Hippo	34.2	37.6	Control (Path) 4 Temporal Ctx	40.6	65.5
AD 3 Hippo	12.0	15.8	AD 1 Occipital Ctx	24.7	29.1
AD 4 Hippo	13.8	10.3	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	60.7	57.8	AD 3 Occipital Ctx	14.7	15.0
AD 6 Hippo	80.7	72.2	AD 4 Occipital Ctx	35.4	22.4
Control 2 Hippo	35.8	38.4	AD 5 Occipital Ctx	3.9	30.4
Control 4 Hippo	16.5	11.7	AD 6 Occipital	46.0	37.4

			Ctx		
Control (Path) 3 Hippo	13.1	15.4	Control 1 Occipital Ctx	9.9	10.7
AD 1 Temporal Ctx	39.0	31.4	Control 2 Occipital Ctx	39.0	38.4
AD 2 Temporal Ctx	38.7	73.2	Control 3 Occipital Ctx	23.0	20.6
AD 3 Temporal Ctx	9.5	13.2	Control 4 Occipital Ctx	13.3	13.3
AD 4 Temporal Ctx	27.9	34.9	Control (Path) 1 Occipital Ctx	80.1	76.3
AD 5 Inf Temporal Ctx	59.0	100.0	Control (Path) 2 Occipital Ctx	17.3	20.0
AD 5 SupTemporal Ctx	33.2	44.1	Control (Path) 3 Occipital Ctx	8.4	8.7
AD 6 Inf Temporal Ctx	100.0	73.2	Control (Path) 4 Occipital Ctx	21.2	20.6
AD 6 Sup Temporal Ctx	79.6	80.1	Control 1 Parietal Ctx	12.1	16.3
Control 1 Temporal Ctx	10.2	13.7	Control 2 Parietal Ctx	48.0	40.9
Control 2 Temporal Ctx	41.2	31.9	Control 3 Parietal Ctx	17.9	16.3
Control 3 Temporal Ctx	20.3	20.0	Control (Path) 1 Parietal Ctx	74.7	64.2
Control 4 Temporal Ctx	9.7	9.9	Control (Path) 2 Parietal Ctx	28.9	59.9
Control (Path) 1 Temporal Ctx	59.9	68.3	Control (Path) 3 Parietal Ctx	10.2	9.0
Control (Path) 2 Temporal Ctx	40.3	41.2	Control (Path) 4 Parietal Ctx	44.8	43.8

Table BRD. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119	Tissue Name	Rel. Exp.(%) Ag3373, Run 217043119
Adipose	12.0	Renal ca. TK-10	20.3
Melanoma* Hs688(A).T	30.8	Bladder	23.2
Melanoma* Hs688(B).T	69.3	Gastric ca. (liver met.) NCI-N87	25.3
Melanoma* M14	15.0	Gastric ca. KATO III	30.8
Melanoma* LOXIMVI	26.6	Colon ca. SW-948	9.7
Melanoma* SK- MEL-5	21.5	Colon ca. SW480	35.1
Squamous cell carcinoma SCC-4	33.0	Colon ca.* (SW480 met) SW620	13.9
Testis Pool	19.8	Colon ca. HT29	8.5
Prostate ca.* (bone met) PC-3	100.0	Colon ca. HCT-116	36.9
Prostate Pool	9.2	Colon ca. CaCo-2	42.9
Placenta	3.8	Colon cancer tissue	9.0
Uterus Pool	7.4	Colon ca. SW1116	5.8
Ovarian ca. OVCAR-3	28.5	Colon ca. Colo-205	4.3
Ovarian ca. SK-OV- 3	40.3	Colon ca. SW-48	4.2
Ovarian ca. OVCAR-4	20.0	Colon Pool	20.7
Ovarian ca. OVCAR-5	35.1	Small Intestine Pool	12.2
Ovarian ca. IGROV- 1	10.9	Stomach Pool	9.9
Ovarian ca. OVCAR-8	9.2	Bone Marrow Pool	11.6
Ovary	9.7	Fetal Heart	20.7
Breast ca. MCF-7	37.6	Heart Pool	10.6
Breast ca. MDA- MB-231	37.1	Lymph Node Pool	17.9
Breast ca. BT 549	62.4	Fetal Skeletal Muscle	12.3
Breast ca. T47D	61.1	Skeletal Muscle Pool	16.0
Breast ca. MDA-N	10.0	Spleen Pool	11.6
Breast Pool	17.3	Thymus Pool	12.2
Trachea	12.0	CNS cancer (glio/astro) U87-MG	29.1
Lung	6.7	CNS cancer (glio/astro) U-118-MG	69.3

Fetal Lung	34.2	CNS cancer (neuro;met) SK-N-AS	34.9
Lung ca. NCI-N417	5.4	CNS cancer (astro) SF-539	19.1
Lung ca. LX-1	17.2	CNS cancer (astro) SNB-75	35.8
Lung ca. NCI-H146	3.0	CNS cancer (glio) SNB-19	11.3
Lung ca. SHP-77	18.6	CNS cancer (glio) SF-295	26.4
Lung ca. A549	29.1	Brain (Amygdala) Pool	4.5
Lung ca. NCI-H526	4.6	Brain (cerebellum)	8.1
Lung ca. NCI-H23	31.6	Brain (fetal)	13.2
Lung ca. NCI-H460	18.2	Brain (Hippocampus) Pool	5.3
Lung ca. HOP-62	14.1	Cerebral Cortex Pool	5.4
Lung ca. NCI-H522	31.6	Brain (Substantia nigra) Pool	4.8
Liver	1.2	Brain (Thalamus) Pool	8.0
Fetal Liver	32.3	Brain (whole)	6.2
Liver ca. HepG2	14.6	Spinal Cord Pool	6.6
Kidney Pool	22.1	Adrenal Gland	8.1
Fetal Kidney	26.1	Pituitary gland Pool	3.0
Renal ca. 786-0	28.7	Salivary Gland	4.7
Renal ca. A498	11.3	Thyroid (female)	4.4
Renal ca. ACHN	12.2	Pancreatic ca. CAPAN2	17.3
Renal ca. UO-31	24.1	Pancreas Pool	17.1

Table BRE. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931	Tissue Name	Rel. Exp.(%) Ag3023, Run 167966931
Liver adenocarcinoma	51.1	Kidney (fetal)	26.2
Pancreas	6.1	Renal ca. 786-0	34.2
Pancreatic ca. CAPAN 2	17.7	Renal ca. A498	17.6
Adrenal gland	3.8	Renal ca. RXF 393	17.2
Thyroid	3.0	Renal ca. ACHN	13.5
Salivary gland	3.9	Renal ca. UO-31	0.0
Pituitary gland	3.6	Renal ca. TK-10	23.0
Brain (fetal)	8.1	Liver	11.7
Brain (whole)	8.5	Liver (fetal)	8.0
Brain (amygdala)	6.7	Liver ca. (hepatoblast) HepG2	26.2

Brain (cerebellum)	15.2	Lung	3.1
Brain (hippocampus)	5.4	Lung (fetal)	11.0
Brain (substantia nigra)	9.0	Lung ca. (small cell) LX-1	12.9
Brain (thalamus)	4.2	Lung ca. (small cell) NCI-H69	9.9
Cerebral Cortex	2.0	Lung ca. (s.cell var.) SHP-77	67.8
Spinal cord	6.9	Lung ca. (large cell)NCI-H460	3.4
glio/astro U87-MG	28.5	Lung ca. (non-sm. cell) A549	45.1
glio/astro U-118-MG	46.7	Lung ca. (non-s.cell) NCI-H23	22.7
astrocytoma SW1783	40.6	Lung ca. (non-s.cell) HOP-62	25.7
neuro*; met SK-N-AS	27.2	Lung ca. (non-s.cl) NCI-H522	38.2
astrocytoma SF-539	29.7	Lung ca. (squam.) SW 900	27.4
astrocytoma SNB-75	35.1	Lung ca. (squam.) NCI-H596	29.9
glioma SNB-19	15.6	Mammary gland	5.1
glioma U251	37.9	Breast ca.* (pl.ef) MCF-7	47.0
glioma SF-295	18.4	Breast ca.* (pl.ef) MDA-MB-231	22.7
Heart (fetal)	2.9	Breast ca.* (pl.ef) T47D	86.5
Heart	12.9	Breast ca. BT-549	15.9
Skeletal muscle (fetal)	3.4	Breast ca. MDA-N	10.4
Skeletal muscle	36.3	Ovary	2.9
Bone marrow	4.5	Ovarian ca. OVCAR- 3	26.1
Thymus	14.3	Ovarian ca. OVCAR- 4	16.3
Spleen	8.7	Ovarian ca. OVCAR- 5	83.5
Lymph node	11.8	Ovarian ca. OVCAR- 8	9.3
Colorectal	10.4	Ovarian ca. IGROV- 1	12.0
Stomach	7.8	Ovarian ca.* (ascites) SK-OV-3	100.0
Small intestine	5.1	Uterus	4.9

Colon ca. SW480	19.3	Placenta	1.3
Colon ca.* SW620(SW480 met)	42.9	Prostate	3.9
Colon ca. HT29	9.9	Prostate ca.* (bone met)PC-3	78.5
Colon ca. HCT-116	26.2	Testis	9.7
Colon ca. CaCo-2	41.5	Melanoma Hs688(A).T	5.9
Colon ca. tissue(ODO3866)	6.3	Melanoma* (met) Hs688(B).T	14.2
Colon ca. HCC-2998	16.0	Melanoma UACC-62	14.0
Gastric ca.* (liver met) NCI-N87	18.8	Melanoma M14	5.7
Bladder	30.6	Melanoma LOX IMVI	8.8
Trachea	3.2	Melanoma* (met) SK-MEL-5	14.7
Kidney	9.6	Adipose	18.9

Table BRF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617	Tissue Name	Rel. Exp.(%) Ag3023, Run 164516146	Rel. Exp.(%) Ag3373, Run 165296617
Secondary Th1 act	18.6	17.9	HUVEC IL-1beta	20.3	18.6
Secondary Th2 act	24.3	28.5	HUVEC IFN gamma	25.3	22.7
Secondary Tr1 act	22.8	21.8	HUVEC TNF alpha + IFN gamma	16.3	18.0
Secondary Th1 rest	7.5	6.8	HUVEC TNF alpha + IL4	18.2	13.4
Secondary Th2 rest	11.6	9.5	HUVEC IL-11	13.7	9.9
Secondary Tr1 rest	12.1	10.7	Lung Microvascular EC none	25.7	21.6
Primary Th1 act	20.7	16.5	Lung Microvascular EC TNFalpha + IL- 1beta	26.2	18.3
Primary Th2 act	20.2	19.3	Microvascular Dermal EC none	27.5	21.3
Primary Tr1 act	23.3	27.7	Microvascular Dermal EC	20.7	19.9

			TNFalpha + IL-1beta		
Primary Th1 rest	51.1	51.4	Bronchial epithelium TNFalpha + IL1beta	13.0	16.3
Primary Th2 rest	26.2	29.5	Small airway epithelium none	8.1	8.5
Primary Tr1 rest	23.7	26.1	Small airway epithelium TNFalpha + IL-1beta	50.3	39.8
CD45RA CD4 lymphocyte act	14.6	11.0	Coronary artery SMC rest	20.2	18.9
CD45RO CD4 lymphocyte act	25.2	22.4	Coronary artery SMC TNFalpha + IL-1beta	12.0	9.8
CD8 lymphocyte act	20.4	15.8	Astrocytes rest	10.4	11.1
Secondary CD8 lymphocyte rest	16.5	19.9	Astrocytes TNFalpha + IL-1beta	11.7	9.8
Secondary CD8 lymphocyte act	13.2	9.3	KU-812 (Basophil) rest	47.6	38.2
CD4 lymphocyte none	17.1	11.6	KU-812 (Basophil) PMA/ionomycin	94.0	92.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	18.3	16.6	CCD1106 (Keratinocytes) none	19.9	13.2
LAK cells rest	25.5	16.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.0	4.8
LAK cells IL-2	27.2	22.5	Liver cirrhosis	3.1	2.7
LAK cells IL-2+IL-12	27.2	19.3	Lupus kidney	2.1	1.7
LAK cells IL-2+IFN gamma	36.3	34.4	NCI-H292 none	30.1	18.9
LAK cells IL-2+IL-18	35.1	29.7	NCI-H292 IL-4	33.9	34.6
LAK cells PMA/ionomycin	12.4	11.0	NCI-H292 IL-9	40.1	29.1
NK Cells IL-2 rest	20.0	15.0	NCI-H292 IL-13	16.2	14.2
Two Way MLR 3 day	24.0	16.7	NCI-H292 IFN gamma	16.6	18.4
Two Way MLR 5	12.9	10.1	HPAEC none	13.6	13.5

day					
Two Way MLR 7 day	11.4	9.5	HPAEC TNF alpha + IL-1 beta	25.3	25.3
PBMC rest	13.7	10.5	Lung fibroblast none	11.4	14.2
PBMC PWM	69.3	66.4	Lung fibroblast TNF alpha + IL-1 beta	6.1	7.2
PBMC PHA-L	22.8	17.7	Lung fibroblast IL-4	28.5	29.1
Ramos (B cell) none	24.1	19.3	Lung fibroblast IL-9	23.0	23.3
Ramos (B cell) ionomycin	100.0	100.0	Lung fibroblast IL-13	20.6	18.9
B lymphocytes PWM	71.7	74.2	Lung fibroblast IFN gamma	39.0	32.5
B lymphocytes CD40L and IL-4	29.1	28.7	Dermal fibroblast CCD1070 rest	33.9	31.0
EOL-1 dbcAMP	12.1	10.5	Dermal fibroblast CCD1070 TNF alpha	76.8	62.0
EOL-1 dbcAMP PMA/ionomycin	14.5	10.9	Dermal fibroblast CCD1070 IL-1 beta	20.3	13.9
Dendritic cells none	13.2	14.8	Dermal fibroblast IFN gamma	14.2	9.5
Dendritic cells LPS	11.7	8.3	Dermal fibroblast IL-4	26.4	20.4
Dendritic cells anti-CD40	17.7	12.7	IBD Colitis 2	2.6	2.2
Monocytes rest	16.7	17.6	IBD Crohn's	2.0	1.9
Monocytes LPS	6.4	5.0	Colon	11.9	10.5
Macrophages rest	23.5	22.8	Lung	13.3	11.2
Macrophages LPS	9.9	7.1	Thymus	14.4	12.9
HUVEC none	20.6	17.9	Kidney	27.5	19.6
HUVEC starved	43.5	38.4			

5 **CNS_neurodegeneration_v1.0 Summary:** Ag3023/Ag3373 This panel does not show differential expression of the NOV77 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.4 Summary: Ag3373 Highest expression of the NOV77 gene is seen in a prostate cancer cell line (CT=27). Overall, this gene is expressed at

moderate levels in the cancer cell lines in this panel. A higher level of expression is observed in clusters of cell lines derived from prostate, brain, melanoma, colon, lung, breast and ovarian cancer when compared to expression in normal prostate, brain, colon, lung, breast and ovary.

Thus, this gene could potentially be used as a diagnostic marker of cancer in these tissues.

5 Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, this gene product has moderate levels of expression in adipose, heart, skeletal muscle, adrenal, pituitary, thyroid and pancreas. Thus, this gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes.

10 In addition, this gene appears to be differentially expressed in fetal (CT value = 29) vs adult liver (CT value = 33) and may be useful for differentiation between the two sources of this tissue.

This gene is also expressed at moderate levels in all central nervous system samples present on this panel. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3023 The NOV77 gene is ubiquitously expressed among the samples on this panel, with highest expression in an ovarian cancer cell line (CT=28.8). Overall, the expression of this gene shows good agreement with panel 1.4. A higher level of expression is observed in prostate, brain, melanoma, colon, lung, pancreatic, breast and ovarian cancer cell lines than the normal prostate, brain, colon, lung, pancreas, breast and ovary. Thus, expression of this gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

25 Among tissues with metabolic function, expression of this gene is widespread, as in the previous panel. Please see Panel 1.4 for discussion of utility of this gene in metabolic disease.

This gene represents a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes, members of which are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

References:

Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. *Brain Res Mol Brain Res* 1995 Feb;28(2):353-6

PAC-1 mRNA has previously been found only in activated T-cells in vitro and in vivo.

5 The gene encodes a dual specificity protein phosphatase that regulates MAP kinase activity. Here, I describe that PAC-1 mRNA is induced also in neurons in the rat brain following 30 min of forebrain ischemia. At 6, 12 and 24 h after ischemia, PAC-1 mRNA was found most prominently in hippocampal cells which are resistant to 30 min of forebrain ischemia, but not in the selectively vulnerable CA1 sector. At later time points and in control animals no PAC-1

10 mRNA could be detected in any brain region. The protein-tyrosine/threonine phosphatase PAC-1, therefore, may be involved in adaptational responses of hippocampal cells resistant to ischemic injury.

Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. *Neuroreport* 1997 Sep

15 29;8(14):3077-80

Recurrent seizure activity leads to delayed neuronal death as well as to inflammatory responses involving microglia in hippocampal subfields CA1, CA3 and CA4. Since mitogen activated protein (MAP) kinases control neuronal apoptosis and trigger generation of inflammatory cytokines, their activation state could determine seizure-related brain damage.

20 PAC1 is a dual specificity protein phosphatase inactivating MAP kinases which we have found to be undetectable in normal brain. Despite this, kainic acid-induced seizure activity lead to rapid (approximately 3 h) but transient appearance of PAC1 mRNA in granule cells of the dentate gyrus as well as in pyramidal CA1 neurons. This pattern changed with time and after 2-3 days PAC1 was induced in dying CA1 and CA3 neurons. At this time PAC1 mRNA

25 was also expressed in white matter microglia as well as in microglia invading the damaged hippocampus. PAC1 may play an important role controlling MAP kinase involvement in both neuronal death and neuro-inflammation following excitotoxic damage.

Panel 4D Summary: Ag3023/Ag3373 The NOV77 gene is expressed at high to moderate levels in a wide range of cell types and tissues of significance in the immune

30 response in health and disease. Highest expression of this gene is seen in ionomycin treated Ramos B cells (CT=26.83). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering

from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

NOV78

- 5 Expression of gene NOV78 was assessed using the primer-probe set Ag3025, described in Table BSA. Results of the RTQ-PCR runs are shown in Tables BSB, BSC and BSD.

Table BSA. Probe Name Ag3025

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gctgctgcaattgtaataggtt-3'	22	596	1309
Probe	TET-5'-tcttgatgaattctgaacaaacctca-3'-TAMRA	26	618	1310
Reverse	5'-catatggaaggctcttgcatttt-3'	22	669	1311

Table BSB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3025, Run 209821733	Tissue Name	Rel. Exp.(%) Ag3025, Run 209821733
AD 1 Hippo	21.3	Control (Path) 3 Temporal Ctx	4.8
AD 2 Hippo	29.1	Control (Path) 4 Temporal Ctx	50.7
AD 3 Hippo	15.9	AD 1 Occipital Ctx	22.7
AD 4 Hippo	11.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	66.4	AD 3 Occipital Ctx	9.0
AD 6 Hippo	26.6	AD 4 Occipital Ctx	13.7
Control 2 Hippo	32.5	AD 5 Occipital Ctx	54.7
Control 4 Hippo	7.2	AD 6 Occipital Ctx	8.1
Control (Path) 3 Hippo	4.8	Control 1 Occipital Ctx	2.3
AD 1 Temporal Ctx	19.9	Control 2 Occipital Ctx	40.1
AD 2 Temporal Ctx	29.9	Control 3 Occipital Ctx	13.8
AD 3 Temporal Ctx	9.3	Control 4 Occipital Ctx	5.6
AD 4 Temporal Ctx	25.2	Control (Path) 1 Occipital Ctx	84.7
AD 5 Inf Temporal Ctx	79.6	Control (Path) 2 Occipital Ctx	10.0
AD 5 Sup Temporal	56.6	Control (Path) 3	0.6

Ctx		Occipital Ctx	
AD 6 Inf Temporal Ctx	19.3	Control (Path) 4 Occipital Ctx	12.3
AD 6 Sup Temporal Ctx	24.5	Control 1 Parietal Ctx	3.8
Control 1 Temporal Ctx	9.7	Control 2 Parietal Ctx	30.1
Control 2 Temporal Ctx	43.5	Control 3 Parietal Ctx	21.6
Control 3 Temporal Ctx	15.2	Control (Path) 1 Parietal Ctx	100.0
Control 3 Temporal Ctx	7.4	Control (Path) 2 Parietal Ctx	34.9
Control (Path) 1 Temporal Ctx	75.3	Control (Path) 3 Parietal Ctx	1.5
Control (Path) 2 Temporal Ctx	35.6	Control (Path) 4 Parietal Ctx	52.1

Table BSC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3025, Run 167968622	Tissue Name	Rel. Exp.(%) Ag3025, Run 167968622
Liver adenocarcinoma	3.6	Kidney (fetal)	16.3
Pancreas	6.8	Renal ca. 786-0	6.2
Pancreatic ca. CAPAN 2	2.7	Renal ca. A498	10.9
Adrenal gland	2.3	Renal ca. RXF 393	6.7
Thyroid	7.2	Renal ca. ACHN	3.3
Salivary gland	3.2	Renal ca. UO-31	5.3
Pituitary gland	16.3	Renal ca. TK-10	7.2
Brain (fetal)	37.6	Liver	3.8
Brain (whole)	33.2	Liver (fetal)	3.4
Brain (amygdala)	29.7	Liver ca. (hepatoblast) HepG2	2.5
Brain (cerebellum)	9.7	Lung	4.6
Brain (hippocampus)	21.9	Lung (fetal)	0.0
Brain (substantia nigra)	12.2	Lung ca. (small cell) LX-1	3.8
Brain (thalamus)	10.4	Lung ca. (small cell) NCI-H69	2.5
Cerebral Cortex	16.3	Lung ca. (s.cell var.) SHP-77	100.0
Spinal cord	10.7	Lung ca. (large cell) NCI-H460	0.8
glio/astro U87-MG	6.0	Lung ca. (non-sm.	5.8

		cell) A549	
glio/astro U-118-MG	8.8	Lung ca. (non-s.cell) NCI-H23	15.4
astrocytoma SW1783	15.3	Lung ca. (non-s.cell) HOP-62	7.7
neuro*; met SK-N-AS	6.4	Lung ca. (non-s.cl) NCI-H522	13.4
astrocytoma SF-539	5.3	Lung ca. (squam.) SW 900	1.7
astrocytoma SNB-75	7.1	Lung ca. (squam.) NCI-H596	3.7
glioma SNB-19	6.3	Mammary gland	4.1
glioma U251	9.9	Breast ca.* (pl.ef) MCF-7	8.7
glioma SF-295	10.4	Breast ca.* (pl.ef) MDA-MB-231	1.7
Heart (fetal)	3.3	Breast ca.* (pl.ef) T47D	14.1
Heart	20.3	Breast ca. BT-549	3.6
Skeletal muscle (fetal)	0.5	Breast ca. MDA-N	15.8
Skeletal muscle	15.5	Ovary	9.0
Bone marrow	2.8	Ovarian ca. OVCAR- 3	3.5
Thymus	11.9	Ovarian ca. OVCAR- 4	1.0
Spleen	5.6	Ovarian ca. OVCAR- 5	19.2
Lymph node	3.1	Ovarian ca. OVCAR- 8	5.1
Colorectal	4.9	Ovarian ca. IGROV- 1	0.6
Stomach	11.9	Ovarian ca.* (ascites) SK-OV-3	52.1
Small intestine	6.2	Uterus	7.3
Colon ca. SW480	0.3	Placenta	3.1
Colon ca.* SW620(SW480 met)	17.0	Prostate	1.4
Colon ca. HT29	2.8	Prostate ca.* (bone met)PC-3	11.6
Colon ca. HCT-116	5.1	Testis	5.4
Colon ca. CaCo-2	15.2	Melanoma Hs688(A).T	2.2
Colon ca. tissue(ODO3866)	1.7	Melanoma* (met) Hs688(B).T	2.2
Colon ca. HCC-2998	9.0	Melanoma UACC-62	12.2

Gastric ca.* (liver met) NCI-N87	3.2	Melanoma M14	3.6
Bladder	3.1	Melanoma LOX IMVI	3.5
Trachea	5.2	Melanoma* (met) SK-MEL-5	11.8
Kidney	17.4	Adipose	7.9

Table BSD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3025, Run 164528140	Tissue Name	Rel. Exp.(%) Ag3025, Run 164528140
Secondary Th1 act	2.6	HUVEC IL-1beta	12.5
Secondary Th2 act	3.3	HUVEC IFN gamma	9.5
Secondary Tr1 act	4.8	HUVEC TNF alpha + IFN gamma	5.1
Secondary Th1 rest	1.2	HUVEC TNF alpha + IL4	5.1
Secondary Th2 rest	2.7	HUVEC IL-11	3.4
Secondary Tr1 rest	1.4	Lung Microvascular EC none	10.7
Primary Th1 act	6.2	Lung Microvascular EC TNFalpha + IL-1beta	13.5
Primary Th2 act	3.5	Microvascular Dermal EC none	6.0
Primary Tr1 act	2.5	Microvascular Dermal EC TNFalpha + IL-1beta	2.6
Primary Th1 rest	5.4	Bronchial epithelium TNFalpha + IL1beta	6.2
Primary Th2 rest	6.7	Small airway epithelium none	0.5
Primary Tr1 rest	1.4	Small airway epithelium TNFalpha + IL-1beta	8.0
CD45RA CD4 lymphocyte act	4.4	Coronary artery SMC rest	32.5
CD45RO CD4 lymphocyte act	4.0	Coronary artery SMC TNFalpha + IL-1beta	6.9
CD8 lymphocyte act	16.8	Astrocytes rest	9.2
Secondary CD8 lymphocyte rest	5.2	Astrocytes TNFalpha + IL-1beta	3.4
Secondary CD8 lymphocyte act	1.9	KU-812 (Basophil) rest	7.3
CD4 lymphocyte none	2.3	KU-812 (Basophil) PMA/ionomycin	29.5
2ry Th1/Th2/Tr1 _anti-	2.6	CCD1106 (Keratinocytes)	1.3

CD95 CH11		none	
LAK cells rest	5.8	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	21.0	Liver cirrhosis	3.7
LAK cells IL-2+IL-12	11.3	Lupus kidney	3.7
LAK cells IL-2+IFN gamma	12.4	NCI-H292 none	12.2
LAK cells IL-2+ IL-18	67.4	NCI-H292 IL-4	10.2
LAK cells PMA/ionomycin	0.6	NCI-H292 IL-9	100.0
NK Cells IL-2 rest	4.2	NCI-H292 IL-13	5.1
Two Way MLR 3 day	3.5	NCI-H292 IFN gamma	2.7
Two Way MLR 5 day	1.3	HPAEC none	9.3
Two Way MLR 7 day	4.9	HPAEC TNF alpha + IL-1 beta	2.4
PBMC rest	4.2	Lung fibroblast none	5.5
PBMC PWM	24.7	Lung fibroblast TNF alpha + IL-1 beta	6.3
PBMC PHA-L	4.3	Lung fibroblast IL-4	6.8
Ramos (B cell) none	19.1	Lung fibroblast IL-9	10.4
Ramos (B cell) ionomycin	33.4	Lung fibroblast IL-13	4.5
B lymphocytes PWM	19.2	Lung fibroblast IFN gamma	15.2
B lymphocytes CD40L and IL-4	5.2	Dermal fibroblast CCD1070 rest	17.8
EOL-1 dbcAMP	0.7	Dermal fibroblast CCD1070 TNF alpha	36.6
EOL-1 dbcAMP PMA/ionomycin	0.9	Dermal fibroblast CCD1070 IL-1 beta	6.5
Dendritic cells none	3.7	Dermal fibroblast IFN gamma	0.6
Dendritic cells LPS	6.0	Dermal fibroblast IL-4	6.6
Dendritic cells anti- CD40	3.3	IBD Colitis 2	0.7
Monocytes rest	2.4	IBD Crohn's	0.8
Monocytes LPS	0.6	Colon	21.5
Macrophages rest	5.6	Lung	18.3
Macrophages LPS	0.8	Thymus	52.1
HUVEC none	6.6	Kidney	19.2
HUVEC starved	21.9		

CNS_neurodegeneration_v1.0 Summary: Ag3025 This panel does not show differential expression of the NOV78 gene in Alzheimer's disease. However, this expression

profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3025 Highest expression of the NOV78 gene is seen in a lung cancer cell line (CT=30.5). Higher levels of expression are observed in prostate, lung, breast and ovarian cancer cell lines when compared with the normal prostate, lung, breast and ovary. Thus, expression of this gene may be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be effective in the treatment of cancer in these tissues.

Among tissues with metabolic function, this gene has a low level of expression in pancreas, thyroid, pituitary, heart, and adipose. Therefore, this gene product may be a small molecule target for the treatment of metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

This gene represents a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes that are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

References:

Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. *Brain Res Mol Brain Res* 1995 Feb;28(2):353-6

PAC-1 mRNA has previously been found only in activated T-cells in vitro and in vivo. The gene encodes a dual specificity protein phosphatase that regulates MAP kinase activity. Here, I describe that PAC-1 mRNA is induced also in neurons in the rat brain following 30 min of forebrain ischemia. At 6, 12 and 24 h after ischemia, PAC-1 mRNA was found most prominently in hippocampal cells which are resistant to 30 min of forebrain ischemia, but not in the selectively vulnerable CA1 sector. At later time points and in control animals no PAC-1 mRNA could be detected in any brain region. The protein-tyrosine/threonine phosphatase PAC-1, therefore, may be involved in adaptational responses of hippocampal cells resistant to ischemic injury.

Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. *Neuroreport* 1997 Sep 29;8(14):3077-80

Recurrent seizure activity leads to delayed neuronal death as well as to inflammatory responses involving microglia in hippocampal subfields CA1, CA3 and CA4. Since mitogen activated protein (MAP) kinases control neuronal apoptosis and trigger generation of inflammatory cytokines, their activation state could determine seizure-related brain damage.

5 PAC1 is a dual specificity protein phosphatase inactivating MAP kinases which we have found to be undetectable in normal brain. Despite this, kainic acid-induced seizure activity lead to rapid (approximately 3 h) but transient appearance of PAC1 mRNA in granule cells of the dentate gyrus as well as in pyramidal CA1 neurons. This pattern changed with time and after 2-3 days PAC1 was induced in dying CA1 and CA3 neurons. At this time PAC1 mRNA

10 was also expressed in white matter microglia as well as in microglia invading the damaged hippocampus. PAC1 may play an important role controlling MAP kinase involvement in both neuronal death and neuro-inflammation following excitotoxic damage.

Panel 4D Summary: Ag3025 The NOV78 gene is expressed at moderate to low levels in a wide range of cell types of significance Highest expression is detected in IL-9 treated

15 NCI-H292 mucoepidermoid cells (CT=31.81) with lower expression levels in non-treated NCI-H292 cells. Expression is also seen in (i) LAK cells stimulated with IL-2, IL-2 +IL-12, IL-2 + IL-18, and IL-2 + IFNgamma (ii) stimulated and non-stimulated Ramos B cells and polkweed mitogen stimulated B lymphocytes, (iii) starved and IL-1 treated HUVECs, (iv) TNF alpha+IL-1 beta treated and non treated lung microvascular endothelial cells and resting

20 coronary artery smooth muscle cells (v) treated Ku-812 basophils (vi) IFN gamma treated lung fibroblasts, and (vii) normal tissues represented by colon, lung, thymus and kidney. Based on this pattern of expression, this gene product may be involved in both disease and homeostatic processes for these and other cell types and tissues. Therefore, modulation of this gene product with a functional therapeutic may lead to the alteration of functions associated with these cell

25 and tissue types and improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as COPD, emphysema, asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

NOV79: Dual Specificity Phosphatase

Expression of gene NOV79 was assessed using the primer-probe set Ag3039,

30 described in Table BTA. Results of the RTQ-PCR runs are shown in Tables BTB, BTC and BTD.

Table BTA. Probe Name Ag3039

Primers	Sequences	Length	Start	SEQ ID NO:
---------	-----------	--------	-------	------------

			Position	
Forward	5'-gccgaaataagatcacacacat-3'	22	320	1312
Probe	TET-5'-tctatccatgagtcaccccagcctct-3'- TAMRA	26	346	1313
Reverse	5'-atgcgaaggtaggtgatatcct-3'	22	377	1314

Table BTB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3039, Run 211012103	Tissue Name	Rel. Exp.(%) Ag3039, Run 211012103
AD 1 Hippo	18.4	Control (Path) 3 Temporal Ctx	8.2
AD 2 Hippo	48.0	Control (Path) 4 Temporal Ctx	36.3
AD 3 Hippo	9.8	AD 1 Occipital Ctx	9.5
AD 4 Hippo	13.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	70.2	AD 3 Occipital Ctx	6.3
AD 6 Hippo	69.3	AD 4 Occipital Ctx	20.9
Control 2 Hippo	25.5	AD 5 Occipital Ctx	18.3
Control 4 Hippo	24.0	AD 6 Occipital Ctx	43.2
Control (Path) 3 Hippo	7.6	Control 1 Occipital Ctx	6.0
AD 1 Temporal Ctx	24.3	Control 2 Occipital Ctx	57.0
AD 2 Temporal Ctx	36.9	Control 3 Occipital Ctx	18.7
AD 3 Temporal Ctx	4.7	Control 4 Occipital Ctx	13.9
AD 4 Temporal Ctx	24.5	Control (Path) 1 Occipital Ctx	74.2
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	14.8
AD 5 Sup Temporal Ctx	62.9	Control (Path) 3 Occipital Ctx	4.3
AD 6 Inf Temporal Ctx	58.2	Control (Path) 4 Occipital Ctx	25.2
AD 6 Sup Temporal Ctx	49.3	Control 1 Parietal Ctx	15.9
Control 1 Temporal Ctx	11.6	Control 2 Parietal Ctx	58.2
Control 2 Temporal Ctx	34.4	Control 3 Parietal Ctx	32.1
Control 3 Temporal Ctx	20.0	Control (Path) 1 Parietal Ctx	66.9
Control 4 Temporal	20.7	Control (Path) 2	39.0

Ctx		Parietal Ctx	
Control (Path) 1 Temporal Ctx	44.4	Control (Path) 3 Parietal Ctx	4.6
Control (Path) 2 Temporal Ctx	30.4	Control (Path) 4 Parietal Ctx	35.8

Table BTC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3039, Run 167961816	Tissue Name	Rel. Exp.(%) Ag3039, Run 167961816
Liver adenocarcinoma	1.7	Kidney (fetal)	38.2
Pancreas	2.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.4
Adrenal gland	1.2	Renal ca. RXF 393	0.1
Thyroid	6.0	Renal ca. ACHN	1.5
Salivary gland	0.8	Renal ca. UO-31	0.0
Pituitary gland	3.0	Renal ca. TK-10	0.0
Brain (fetal)	7.4	Liver	0.3
Brain (whole)	7.7	Liver (fetal)	0.7
Brain (amygdala)	6.0	Liver ca. (hepatoblast) HepG2	0.1
Brain (cerebellum)	5.9	Lung	1.3
Brain (hippocampus)	3.9	Lung (fetal)	3.0
Brain (substantia nigra)	16.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	3.8	Lung ca. (small cell) NCI-H69	1.5
Cerebral Cortex	10.8	Lung ca. (s.cell var.) SHP-77	3.7
Spinal cord	15.2	Lung ca. (large cell)NCI-H460	0.1
glio/astro U87-MG	0.2	Lung ca. (non-sm. cell) A549	1.6
glio/astro U-118-MG	0.1	Lung ca. (non-s.cell) NCI-H23	0.4
astrocytoma SW1783	0.6	Lung ca. (non-s.cell) HOP-62	0.0
neuro*; met SK-N-AS	0.3	Lung ca. (non-s.cl) NCI-H522	1.4
astrocytoma SF-539	1.5	Lung ca. (squam.) SW 900	0.7
astrocytoma SNB-75	1.2	Lung ca. (squam.) NCI-H596	5.7
glioma SNB-19	2.1	Mammary gland	2.4

glioma U251	0.3	Breast ca.* (pl.ef) MCF-7	0.7
glioma SF-295	2.3	Breast ca.* (pl.ef) MDA-MB-231	0.1
Heart (fetal)	14.5	Breast ca.* (pl.ef) T47D	13.5
Heart	3.4	Breast ca. BT-549	1.2
Skeletal muscle (fetal)	5.1	Breast ca. MDA-N	6.4
Skeletal muscle	0.0	Ovary	3.9
Bone marrow	0.4	Ovarian ca. OVCAR-3	0.3
Thymus	0.2	Ovarian ca. OVCAR-4	13.8
Spleen	3.1	Ovarian ca. OVCAR-5	1.5
Lymph node	0.9	Ovarian ca. OVCAR-8	0.0
Colorectal	0.4	Ovarian ca. IGROV-1	0.0
Stomach	0.6	Ovarian ca.* (ascites) SK-OV-3	0.4
Small intestine	1.0	Uterus	2.2
Colon ca. SW480	6.7	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.0	Prostate	2.7
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	100.0
Colon ca. CaCo-2	12.2	Melanoma Hs688(A).T	0.1
Colon ca. tissue(ODO3866)	1.2	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	3.0	Melanoma UACC-62	5.2
Gastric ca.* (liver met) NCI-N87	1.2	Melanoma M14	0.8
Bladder	1.8	Melanoma LOX IMVI	0.0
Trachea	0.6	Melanoma* (met) SK-MEL-5	2.0
Kidney	41.8	Adipose	0.6

Table BTD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3039, Run	Tissue Name	Rel. Exp.(%) Ag3039, Run
-------------	-----------------------------	-------------	-----------------------------

	162427949		162427949
Secondary Th1 act	0.0	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	0.3
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.3
Secondary Tr1 rest	0.0	Lung Microvascular EC none	1.3
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.0	Microvascular Dermal EC none	0.3
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	0.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	1.4
Primary Th2 rest	0.0	Small airway epithelium none	1.1
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.3
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	4.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	2.3
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.4
LAK cells rest	0.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.4
LAK cells IL-2	0.4	Liver cirrhosis	0.7
LAK cells IL-2+IL-12	0.0	Lupus kidney	5.2
LAK cells IL-2+IFN gamma	0.4	NCI-H292 none	3.5
LAK cells IL-2+ IL-18	0.8	NCI-H292 IL-4	0.9
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	1.8
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	1.2
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	3.6

Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.3	Lung fibroblast none	0.0
PBMC PWM	2.5	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	1.6	Lung fibroblast IL-4	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	0.5
B lymphocytes PWM	5.5	Lung fibroblast IFN gamma	0.0
B lymphocytes CD40L and IL-4	1.3	Dermal fibroblast CCD1070 rest	0.0
EOL-1 dbcAMP	0.6	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
Dendritic cells none	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells LPS	0.4	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.0	IBD Colitis 2	0.0
Monocytes rest	0.0	IBD Crohn's	1.1
Monocytes LPS	0.4	Colon	2.5
Macrophages rest	1.0	Lung	5.1
Macrophages LPS	0.4	Thymus	100.0
HUVEC none	0.6	Kidney	3.1
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3039 No differential expression of the NOV79 gene is detected between the postmortem brains of Alzheimer's diseased patients and those of non-demented controls. However, this panel confirms the expression of this gene in the CNS. Please see panel 1.3D for a discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3039 Highest expression of the NOV79 gene is seen in the testis (CT=29). In addition, expression of this gene is extremely low in renal and brain cancer cell lines but is expressed in the normal brain and kidney tissues on this sample. Therefore, this gene may be used as a diagnostic marker for brain and kidney cancer and prostate tissue. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of brain and renal cancers.

In addition, this gene is expressed at low levels in metabolic tissues including pancreas, adrenal, thyroid, pituitary, adult and fetal heart, and adipose. This novel protein phosphatase may be a small molecule target for the treatment of metabolic and endocrine disease, including obesity and Types 1 and 2 diabetes. This gene is also differentially expressed in fetal (CT values = 32-33) vs adult skeletal muscle (CT values = 35-40) and may be useful for the differentiation of adult and fetal skeletal muscle.

This gene represents a dual specificity phosphatase that is also expressed at low to moderate levels across the CNS. Dual-specificity phosphatases comprise a family of MAP kinase regulating enzymes that are upregulated in brains subjected to insults such as ischemia and seizure activity. MAP kinases are known to regulate neurotrophic and neurotoxic pathways. Consequently, agents that modulate the activity of this gene may have utility in attenuating the apoptotic and neurodegenerative processes following brain insults.

References:

Wiessner C. The dual specificity phosphatase PAC-1 is transcriptionally induced in the rat brain following transient forebrain ischemia. *Brain Res Mol Brain Res* 1995 Feb;28(2):353-6

PAC-1 mRNA has previously been found only in activated T-cells in vitro and in vivo. The gene encodes a dual specificity protein phosphatase that regulates MAP kinase activity. Here, I describe that PAC-1 mRNA is induced also in neurons in the rat brain following 30 min of forebrain ischemia. At 6, 12 and 24 h after ischemia, PAC-1 mRNA was found most prominently in hippocampal cells which are resistant to 30 min of forebrain ischemia, but not in the selectively vulnerable CA1 sector. At later time points and in control animals no PAC-1 mRNA could be detected in any brain region. The protein-tyrosine/threonine phosphatase PAC-1, therefore, may be involved in adaptational responses of hippocampal cells resistant to ischemic injury.

Boschert U, Muda M, Camps M, Dickinson R, Arkinstall S. Induction of the dual specificity phosphatase PAC1 in rat brain following seizure activity. *Neuroreport* 1997 Sep 29;8(14):3077-80

Recurrent seizure activity leads to delayed neuronal death as well as to inflammatory responses involving microglia in hippocampal subfields CA1, CA3 and CA4. Since mitogen activated protein (MAP) kinases control neuronal apoptosis and trigger generation of inflammatory cytokines, their activation state could determine seizure-related brain damage. PAC1 is a dual specificity protein phosphatase inactivating MAP kinases which we have found to be undetectable in normal brain. Despite this, kainic acid-induced seizure activity

lead to rapid (approximately 3 h) but transient appearance' of PAC1 mRNA in granule cells of the dentate gyrus as well as in pyramidal CA1 neurons. This pattern changed with time and after 2-3 days PAC1 was induced in dying CA1 and CA3 neurons. At this time PAC1 mRNA was also expressed in white matter microglia as well as in microglia invading the damaged hippocampus. PAC1 may play an important role controlling MAP kinase involvement in both neuronal death and neuro-inflammation following excitotoxic damage.

Panel 4D Summary: Ag3039 Expression of the NOV79 gene is highest and almost exclusive to the thymus (CTs=29-30). Expression of this gene could be used to distinguish thymus from the other samples on this panel. The putative dual-specificity phosphatase encoded by this gene may play an important role in T cell development. Small molecule therapeutics designed against the protein encoded by this gene could therefore be utilized to modulate immune function (T cell development) and be important for organ transplant, AIDS treatment or post chemotherapy immune reconstitution.

NOV80

Expression of gene NOV80 was assessed using the primer-probe set Ag3044, described in Table BUA. Results of the RTQ-PCR runs are shown in Tables BUB and BUC.

Table BUA. Probe Name Ag3044

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-tgacgcagaatggaataagct-3'	21	650	1315
Probe	TET-5'-acgtcctctatgccagcaactcctg-3'-TAMRA	25	671	1316
Reverse	5'-gcaagaagtggctctggtagat-3'	22	712	1317

Table BUB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3044, Run 167972762	Tissue Name	Rel. Exp.(%) Ag3044, Run 167972762
Liver adenocarcinoma	0.1	Kidney (fetal)	0.5
Pancreas	0.1	Renal ca. 786-0	0.2
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.1
Thyroid	0.0	Renal ca. ACHN	0.4
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.1	Renal ca. TK-10	0.0
Brain (fetal)	0.4	Liver	0.0
Brain (whole)	0.2	Liver (fetal)	0.0

1364

		SK-OV-3	
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.2	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.1
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.1	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.1	Melanoma M14	0.0
Bladder	0.1	Melanoma LOX IMVI	0.0
Trachea	100.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.2	Adipose	0.0

Table BUC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3044, Run 164334372	Tissue Name	Rel. Exp.(%) Ag3044, Run 164334372
Secondary Th1 act	0.0	HUVEC IL-1beta	3.5
Secondary Th2 act	5.1	HUVEC IFN gamma	0.0
Secondary Tr1 act	1.3	HUVEC TNF alpha + IFN gamma	2.0
Secondary Th1 rest	1.2	HUVEC TNF alpha + IL4	2.0
Secondary Th2 rest	0.7	HUVEC IL-11	0.0
Secondary Tr1 rest	2.1	Lung Microvascular EC none	1.6
Primary Th1 act	2.0	Lung Microvascular EC TNFalpha + IL-1beta	0.3
Primary Th2 act	4.0	Microvascular Dermal EC none	1.6
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	2.7
Primary Th1 rest	9.5	Bronchial epithelium TNFalpha + IL1beta	6.7
Primary Th2 rest	11.9	Small airway epithelium none	2.5
Primary Tr1 rest	4.6	Small airway epithelium	20.9

		TNFalpha + IL-1beta	
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	4.3
CD45RO CD4 lymphocyte act	1.5	Coronary artery SMC TNFalpha + IL-1beta	0.6
CD8 lymphocyte act	5.3	Astrocytes rest	4.3
Secondary CD8 lymphocyte rest	1.9	Astrocytes TNFalpha + IL-1beta	2.9
Secondary CD8 lymphocyte act	22.8	KU-812 (Basophil) rest	6.8
CD4 lymphocyte none	0.1	KU-812 (Basophil) PMA/ionomycin	9.5
2ry Th1/Th2/Tr1_anti-CD95 CH11	2.5	CCD1106 (Keratinocytes) none	2.1
LAK cells rest	4.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	17.7
LAK cells IL-2+IL-12	2.9	Lupus kidney	8.1
LAK cells IL-2+IFN gamma	8.6	NCI-H292 none	52.5
LAK cells IL-2+ IL-18	6.5	NCI-H292 IL-4	20.6
LAK cells PMA/ionomycin	0.2	NCI-H292 IL-9	31.4
NK Cells IL-2 rest	2.1	NCI-H292 IL-13	16.3
Two Way MLR 3 day	1.8	NCI-H292 IFN gamma	11.9
Two Way MLR 5 day	0.0	HPAEC none	0.0
Two Way MLR 7 day	0.0	HPAEC TNF alpha + IL-1 beta	1.0
PBMC rest	4.6	Lung fibroblast none	3.8
PBMC PWM	3.2	Lung fibroblast TNF alpha + IL-1 beta	3.6
PBMC PHA-L	3.0	Lung fibroblast IL-4	5.6
Ramos (B cell) none	0.0	Lung fibroblast IL-9	6.9
Ramos (B cell) ionomycin	0.2	Lung fibroblast IL-13	2.9
B lymphocytes PWM	11.7	Lung fibroblast IFN gamma	3.0
B lymphocytes CD40L and IL-4	86.5	Dermal fibroblast CCD1070 rest	5.8
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	9.9
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.9
Dendritic cells none	4.1	Dermal fibroblast IFN gamma	1.0

Dendritic cells LPS	1.1	Dermal fibroblast IL-4	4.6
Dendritic cells anti-CD40	1.1	IBD Colitis 2	0.0
Monocytes rest	1.7	IBD Crohn's	3.8
Monocytes LPS	2.5	Colon	17.9
Macrophages rest	4.1	Lung	7.6
Macrophages LPS	0.0	Thymus	100.0
HUVEC none	4.8	Kidney	10.1
HUVEC starved	2.2		

Panel 1.3D Summary: Ag3044 Results from one experiment with the NOV80 gene are not included. The amp plot indicates that there were experimental difficulties with this run (data not shown).

5 **Panel 4D Summary:** Ag3044 The NOV80 gene is expressed at low levels in a wide range of cell types of significance in the immune response in health and disease. These cells include: (i) resting LAK and LAK cells stimulated with IL-2+IL-12, IL-2 + IL-18, and IL-2 + IFNgamma (ii) activated primary and secondary Th2 cells, resting primary Th1, Th2 and Tr1 cells, and activated CD8 and secondary CD8 lymphocytes, (iii) IL-1 beta treated HUVECs, 10 (iv) polkweed mitogen stimulated and CD40L + IL-4 stimulated B lymphocytes, (v) treated and non-treated Ku-812 basophils and non-treated dendritic cells, (vi) treated and non-treated peripheral blood mononuclear cells and resting macrophages (vii) treated and non-treated NCI-H292 mucoepidermoid, (viii) treated and non-treated lung fibroblasts, (ix) treated and non-treated astrocytes (x) resting coronary artery SMCs, (xi) resting and TNFalpha treated 15 CCD1070 dermal fibroblasts and IL-4 treated dermal fibroblasts (xi) IBD Crohn's diseases tissue and normal tissues represented by colon, lung, thymus and kidney with the highest expression being detected in thymus tissue (CT=29.81). This expression profile suggests that this gene product may be involved in both disease and homeostatic processes in these and other cell types and tissues. Therefore, modulation of this gene product with a functional 20 therapeutic may lead to the alteration of functions associated with these cell and tissue types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as COPD, emphysema, asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

NOV81a and NOV81b

Expression of gene NOV81a and the full length clone, NOV81b, was assessed using the primer-probe set Ag2906, described in Table BVA. Results of the RTQ-PCR runs are shown in Tables BVB, BVC and BVD.

Table BVA. Probe Name Ag2906

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctacctgggtgaggtctttacc-3'	22	845	1318
Probe	TET-5'-ctccggaagccaggaggaccctt-3'-TAMRA	23	879	1319
Reverse	5'-agaaggactcgggcacatag-3'	20	902	1320

5

Table BVB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2906, Run 162556445	Tissue Name	Rel. Exp.(%) Ag2906, Run 162556445
Liver adenocarcinoma	2.0	Kidney (fetal)	13.5
Pancreas	1.2	Renal ca. 786-0	1.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	2.8
Adrenal gland	1.1	Renal ca. RXF 393	1.9
Thyroid	4.0	Renal ca. ACHN	0.7
Salivary gland	3.3	Renal ca. UO-31	0.0
Pituitary gland	2.4	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	6.8
Brain (amygdala)	2.0	Liver ca. (hepatoblast) HepG2	5.4
Brain (cerebellum)	3.8	Lung	42.0
Brain (hippocampus)	4.7	Lung (fetal)	11.8
Brain (substantia nigra)	0.9	Lung ca. (small cell) LX-1	0.7
Brain (thalamus)	4.6	Lung ca. (small cell) NCI-H69	1.0
Cerebral Cortex	6.4	Lung ca. (s.cell var.) SHP-77	4.5
Spinal cord	7.6	Lung ca. (large cell) NCI-H460	0.0
glio/astro U87-MG	3.0	Lung ca. (non-sm. cell) A549	2.1
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	1.6
astrocytoma SW1783	4.1	Lung ca. (non-s.cell) HOP-62	4.0
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl)	0.0

		NCI-H522	
astrocytoma SF-539	4.5	Lung ca. (squam.) SW 900	8.4
astrocytoma SNB-75	4.9	Lung ca. (squam.) NCI-H596	0.2
glioma SNB-19	7.2	Mammary gland	4.0
glioma U251	3.2	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	4.0	Breast ca.* (pl.ef) MDA-MB-231	3.9
Heart (fetal)	100.0	Breast ca.* (pl.ef) T47D	1.0
Heart	10.7	Breast ca. BT-549	1.8
Skeletal muscle (fetal)	35.8	Breast ca. MDA-N	0.3
Skeletal muscle	0.3	Ovary	2.2
Bone marrow	20.6	Ovarian ca. OVCAR- 3	1.8
Thymus	15.8	Ovarian ca. OVCAR- 4	0.6
Spleen	13.8	Ovarian ca. OVCAR- 5	0.8
Lymph node	5.5	Ovarian ca. OVCAR- 8	0.0
Colorectal	31.2	Ovarian ca. IGROV- 1	0.0
Stomach	7.3	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	31.9	Uterus	4.8
Colon ca. SW480	1.0	Placenta	9.9
Colon ca.* SW620(SW480 met)	0.0	Prostate	9.8
Colon ca. HT29	1.6	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	1.5	Testis	2.4
Colon ca. CaCo-2	4.4	Melanoma Hs688(A).T	1.0
Colon ca. tissue(ODO3866)	8.7	Melanoma* (met) Hs688(B).T	2.6
Colon ca. HCC-2998	8.8	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	2.3	Melanoma M14	0.0
Bladder	9.4	Melanoma LOX IMVI	0.0
Trachea	14.1	Melanoma* (met) SK-MEL-5	2.1

Kidney	11.8	Adipose	6.9
--------	------	---------	-----

Table BVC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2906, Run 162345752	Tissue Name	Rel. Exp.(%) Ag2906, Run 162345752
Normal Colon	39.5	Kidney Margin 8120608	3.0
CC Well to Mod Diff (ODO3866)	16.3	Kidney Cancer 8120613	17.3
CC Margin (ODO3866)	33.9	Kidney Margin 8120614	6.5
CC Gr.2 rectosigmoid (ODO3868)	14.4	Kidney Cancer 9010320	3.4
CC Margin (ODO3868)	1.2	Kidney Margin 9010321	7.2
CC Mod Diff (ODO3920)	27.9	Normal Uterus	4.0
CC Margin (ODO3920)	30.8	Uterus Cancer 064011	4.3
CC Gr.2 ascend colon (ODO3921)	100.0	Normal Thyroid	2.6
CC Margin (ODO3921)	33.9	Thyroid Cancer 064010	3.3
CC from Partial Hepatectomy (ODO4309) Mets	16.2	Thyroid Cancer A302152	10.7
Liver Margin (ODO4309)	2.9	Thyroid Margin A302153	5.5
Colon mets to lung (OD04451-01)	15.6	Normal Breast	7.5
Lung Margin (OD04451- 02)	15.8	Breast Cancer (OD04566)	7.4
Normal Prostate 6546-1	24.5	Breast Cancer (OD04590-01)	11.0
Prostate Cancer (OD04410)	4.8	Breast Cancer Mets (OD04590-03)	30.8
Prostate Margin (OD04410)	5.0	Breast Cancer Metastasis (OD04655-05)	7.0
Prostate Cancer (OD04720-01)	9.5	Breast Cancer 064006	5.7
Prostate Margin (OD04720-02)	19.5	Breast Cancer 1024	7.5
Normal Lung 061010	16.5	Breast Cancer 9100266	13.7
Lung Met to Muscle	3.9	Breast Margin	2.7

(ODO4286)		9100265	
Muscle Margin (ODO4286)	6.2	Breast Cancer A209073	6.1
Lung Malignant Cancer (OD03126)	64.2	Breast Margin A2090734	5.6
Lung Margin (OD03126)	29.1	Normal Liver	1.9
Lung Cancer (OD04404)	16.7	Liver Cancer 064003	0.4
Lung Margin (OD04404)	15.1	Liver Cancer 1025	0.6
Lung Cancer (OD04565)	27.7	Liver Cancer 1026	2.0
Lung Margin (OD04565)	19.8	Liver Cancer 6004-T	0.9
Lung Cancer (OD04237-01)	4.5	Liver Tissue 6004-N	2.8
Lung Margin (OD04237-02)	12.9	Liver Cancer 6005-T	0.4
Ocular Mel Met to Liver (ODO4310)	1.8	Liver Tissue 6005-N	1.0
Liver Margin (ODO4310)	2.6	Normal Bladder	8.5
Melanoma Mets to Lung (OD04321)	4.2	Bladder Cancer 1023	2.4
Lung Margin (OD04321)	35.6	Bladder Cancer A302173	0.8
Normal Kidney	5.4	Bladder Cancer (OD04718-01)	5.8
Kidney Ca, Nuclear grade 2 (OD04338)	4.1	Bladder Normal Adjacent (OD04718-03)	2.0
Kidney Margin (OD04338)	11.0	Normal Ovary	3.1
Kidney Ca Nuclear grade 1/2 (OD04339)	5.0	Ovarian Cancer 064008	11.0
Kidney Margin (OD04339)	2.0	Ovarian Cancer (OD04768-07)	7.9
Kidney Ca, Clear cell type (OD04340)	3.8	Ovary Margin (OD04768-08)	1.3
Kidney Margin (OD04340)	5.6	Normal Stomach	2.9
Kidney Ca, Nuclear grade 3 (OD04348)	2.7	Gastric Cancer 9060358	1.0
Kidney Margin (OD04348)	5.0	Stomach Margin 9060359	3.7
Kidney Cancer (OD04622-01)	5.9	Gastric Cancer 9060395	12.0
Kidney Margin (OD04622-03)	2.1	Stomach Margin 9060394	14.2
Kidney Cancer	0.8	Gastric Cancer	31.9

(OD04450-01)		9060397	
Kidney Margin (OD04450-03)	4.5	Stomach Margin 9060396	5.6
Kidney Cancer 8120607	14.6	Gastric Cancer 064005	8.7

Table BVD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2906, Run 159078634	Tissue Name	Rel. Exp.(%) Ag2906, Run 159078634
Secondary Th1 act	5.6	HUVEC IL-1beta	5.7
Secondary Th2 act	5.3	HUVEC IFN gamma	9.2
Secondary Tr1 act	8.6	HUVEC TNF alpha + IFN gamma	9.5
Secondary Th1 rest	11.0	HUVEC TNF alpha + IL4	4.6
Secondary Th2 rest	6.6	HUVEC IL-11	9.1
Secondary Tr1 rest	10.5	Lung Microvascular EC none	10.0
Primary Th1 act	2.0	Lung Microvascular EC TNFalpha + IL-1beta	3.8
Primary Th2 act	1.2	Microvascular Dermal EC none	10.2
Primary Tr1 act	1.1	Microvascular Dermal EC TNFalpha + IL-1beta	2.5
Primary Th1 rest	10.7	Bronchial epithelium TNFalpha + IL1beta	20.4
Primary Th2 rest	4.4	Small airway epithelium none	8.0
Primary Tr1 rest	18.0	Small airway epithelium TNFalpha + IL-1beta	22.4
CD45RA CD4 lymphocyte act	8.0	Coronary artery SMC rest	4.0
CD45RO CD4 lymphocyte act	5.1	Coronary artery SMC TNFalpha + IL-1beta	6.1
CD8 lymphocyte act	3.7	Astrocytes rest	7.4
Secondary CD8 lymphocyte rest	2.5	Astrocytes TNFalpha + IL-1beta	3.1
Secondary CD8 lymphocyte act	1.5	KU-812 (Basophil) rest	2.6
CD4 lymphocyte none	6.5	KU-812 (Basophil) PMA/ionomycin	2.7
2ry Th1/Th2/Tr1 anti- CD95 CH11	11.6	CCD1106 (Keratinocytes) none	6.4
LAK cells rest	14.6	CCD1106 (Keratinocytes)	3.6

		TNFalpha + IL-1beta	
LAK cells IL-2	8.6	Liver cirrhosis	8.0
LAK cells IL-2+IL-12	6.3	Lupus kidney	2.5
LAK cells IL-2+IFN gamma	14.8	NCI-H292 none	3.7
LAK cells IL-2+ IL-18	2.5	NCI-H292 IL-4	4.9
LAK cells PMA/ionomycin	7.6	NCI-H292 IL-9	0.7
NK Cells IL-2 rest	11.2	NCI-H292 IL-13	1.4
Two Way MLR 3 day	6.4	NCI-H292 IFN gamma	1.3
Two Way MLR 5 day	8.3	HPAEC none	4.0
Two Way MLR 7 day	2.7	HPAEC TNF alpha + IL-1 beta	8.0
PBMC rest	6.2	Lung fibroblast none	5.3
PBMC PWM	8.2	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PHA-L	2.6	Lung fibroblast IL-4	1.6
Ramos (B cell) none	7.5	Lung fibroblast IL-9	2.0
Ramos (B cell) ionomycin	2.2	Lung fibroblast IL-13	3.6
B lymphocytes PWM	2.4	Lung fibroblast IFN gamma	2.7
B lymphocytes CD40L and IL-4	100.0	Dermal fibroblast CCD1070 rest	2.6
EOL-1 dbcAMP	44.4	Dermal fibroblast CCD1070 TNF alpha	8.5
EOL-1 dbcAMP PMA/ionomycin	77.9	Dermal fibroblast CCD1070 IL-1 beta	5.6
Dendritic cells none	21.2	Dermal fibroblast IFN gamma	5.7
Dendritic cells LPS	3.0	Dermal fibroblast IL-4	9.0
Dendritic cells anti-CD40	9.9	IBD Colitis 2	0.0
Monocytes rest	28.1	IBD Crohn's	2.6
Monocytes LPS	10.3	Colon	97.9
Macrophages rest	33.0	Lung	41.5
Macrophages LPS	12.7	Thymus	5.5
HUVEC none	11.9	Kidney	6.9
HUVEC starved	20.4		

Panel 1.3D Summary: Ag2906 The NOV81a gene has a low level of expression in adipose and may be a small molecule target for the treatment of obesity and obesity-related diseases, including Type 2 diabetes. In addition, this gene product appears to be differentially

expressed in fetal (CT value = 31) vs adult heart (CT value = 34) and may be useful for the differentiation between the two tissue types.

Overall, there appears to be higher expression of this gene in the normal tissues compared to the cell lines. Thus, this difference in expression might be of use as a diagnostic marker of cancer.

Panel 2D Summary: Ag2906 The NOV81a gene is expressed at low levels in this panel. A higher level of expression is observed in gastric, bladder, thyroid, breast and ovarian cancer samples when compared to expression in the normal adjacent gastric, bladder, thyroid, breast and ovary tissues. Thus, this gene could potentially be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be useful for the treatment of cancer in these tissues.

Panel 4D Summary: Ag2906 Expression of the NOV81a is widespread in this panel, with highest expression in B lymphocytes treated with CD40L and IL-4 (CT=29.8). Significant expression is also seen in treated eosinophils, resting macrophages and monocytes, and normal colon and lung. Based on this pattern of expression, this gene product may be involved in both disease and homeostatic processes for these and other cell types and tissues. Therefore, modulation of this gene product with a functional therapeutic may lead to the alteration of functions associated with these cell and tissue types and improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as COPD, emphysema, asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis. In addition, the higher levels of expression in colon (CT=30) when compared to colon from patients with inflammatory bowel diseases (IBD)(CTs=35-40) suggests that expression of this gene could be used to differentiate between normal and inflamed colon. Therapeutic modulation of the expression or function of this gene may be effective in the treatment of IBD.

NOV82

Expression of gene NOV82 was assessed using the primer-probe sets Ag3198 and Ag3063, described in Tables BWA and BWB.

Table BWA. Probe Name Ag3198

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5' - cgtgggtcaccagacagttaatt - 3'	22	179	1321
Probe	TET- 5' - cctaccagacaccattgtgtccaagg - 3' - TAMRA	26	212	1322
Reverse	5' - gtcttttcctttgtgcttgtaa - 3'	22	246	1323

Table BWB. Probe Name Ag3063

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cgtggtcaccagacagttaatt-3'	22	179	1324
Probe	TET-5'-cctaccagacaccattgtgtccaagg-3'- TAMRA	26	212	1325
Reverse	5'-gtctttcctttgtgcttgtgaa-3'	22	246	1326

5 **CNS_neurodegeneration_v1.0 Summary:** Ag3198 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a potential probe or chemistry failure.

Panel 1.3D Summary: Ag3063 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a potential probe or chemistry failure.

10 **Panel 4D Summary:** Ag3198 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). The amp plot indicates that there is a high probability of a potential probe or chemistry failure.

NOV83

15 Expression of gene NOV83 was assessed using the primer-probe sets Ag3046 and Ag4125, described in Tables BXA and BXB. Results of the RTQ-PCR runs are shown in Tables BXC and BXD.

Table BXA. Probe Name Ag3046

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaagcaaagaactctgcaagac-3'	22	1215	1327
Probe	TET-5'-ttccagcatgataacttcacagagga-3'- TAMRA	26	1246	1328
Reverse	5'-gagcctgcaaataatcttttgct-3'	22	1272	1329

Table BXB. Probe Name Ag4125

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaagcaaagaactctgcaagac-3'	22	1215	1330
Probe	TET-5'-ttccagcatgataacttcacagagga-3'- TAMRA	26	1246	1331
Reverse	5'-gagcctgcaaataatcttttgct-3'	22	1272	1332

Lung Malignant Cancer (OD03126)	0.0	Breast Margin A2090734	0.0
Lung Margin (OD03126)	0.0	Normal Liver	0.0
Lung Cancer (OD04404)	0.0	Liver Cancer 064003	0.0
Lung Margin (OD04404)	0.0	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	0.0	Liver Cancer 1026	0.0
Lung Margin (OD04565)	0.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	0.0	Liver Tissue 6004-N	0.0
Lung Margin (OD04237-02)	0.0	Liver Cancer 6005-T	0.0
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.0	Normal Bladder	0.0
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	0.0
Lung Margin (OD04321)	0.0	Bladder Cancer A302173	0.0
Normal Kidney	0.0	Bladder Cancer (OD04718-01)	0.0
Kidney Ca, Nuclear grade 2 (OD04338)	0.0	Bladder Normal Adjacent (OD04718-03)	0.0
Kidney Margin (OD04338)	0.0	Normal Ovary	0.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Ovarian Cancer 064008	0.0
Kidney Margin (OD04339)	0.0	Ovarian Cancer (OD04768-07)	0.0
Kidney Ca, Clear cell type (OD04340)	0.0	Ovary Margin (OD04768-08)	0.0
Kidney Margin (OD04340)	0.0	Normal Stomach	0.0
Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer 9060358	0.0
Kidney Margin (OD04348)	0.0	Stomach Margin 9060359	0.0
Kidney Cancer (OD04622-01)	0.0	Gastric Cancer 9060395	0.0
Kidney Margin (OD04622-03)	0.0	Stomach Margin 9060394	0.0
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	0.0
Kidney Margin (OD04450-03)	0.0	Stomach Margin 9060396	0.0

Kidney Cancer 8120607	0.0	Gastric Cancer 064005	0.0
-----------------------	-----	-----------------------	-----

Table BXD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4125, Run 172859315	Tissue Name	Rel. Exp.(%) Ag4125, Run 172859315
Secondary Th1 act	3.0	HUVEC IL-1beta	0.0
Secondary Th2 act	4.2	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	2.7	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	3.0	Microvascular Dermal EC none	0.0
Primary Tr1 act	11.9	Microvascular Dermal EC TNFalpha + IL-1beta	7.4
Primary Th1 rest	7.2	Bronchial epithelium TNFalpha + IL1beta	7.5
Primary Th2 rest	3.1	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	0.0	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	3.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	0.0	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	5.3	NCI-H292 none	0.0

LAK cells IL-2+IFN gamma	1.8	NCI-H292 IL-4	5.8
LAK cells IL-2+ IL-18	3.6	NCI-H292 IL-9	3.5
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	1.3
NK Cells IL-2 rest	5.9	NCI-H292 IFN gamma	0.8
Two Way MLR 3 day	3.3	HPAEC none	0.0
Two Way MLR 5 day	3.2	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.0	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	2.3
PBMC PWM	3.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	1.6	Lung fibroblast IL-9	0.0
Ramos (B cell) none	0.0	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	1.0
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	2.1	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	3.3	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	13.5	Colon	0.0
Macrophages rest	0.0	Lung	5.4
Macrophages LPS	0.0	Thymus	18.4
HUVEC none	0.0	Kidney	100.0
HUVEC starved	0.0		

CNS_neurodegeneration_v1.0 Summary: Ag3046 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Ag4125 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run (data not shown).

General_screening_panel_v1.4 Summary: Ag4125 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag3046 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 2D Summary: Ag3046 Significant expression of this gene is seen exclusively in a breast cancer sample (CT = 25.2). Therefore, expression of this gene may be used to distinguish breast cancers from the other samples on this panel. Furthermore, therapeutic modulation of the activity of the GPCR encoded by this gene may be beneficial in the treatment of breast cancer.

Panel 3D Summary: Ag3046 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4.1D Summary: Ag4125 This gene is only expressed at detectable levels in the kidney (CT = 32.6). The putative GPCR encoded for by this gene could allow cells within the kidney to respond to specific microenvironmental signals (For example, ref. 1). Therefore, antibody or small molecule therapies designed with the protein encoded for by this gene could modulate kidney function and be important in the treatment of inflammatory or autoimmune diseases that affect the kidney, including lupus and glomerulonephritis.

References:

1. Mark M.D., Wittemann S., Herlitz S. (2000) G protein modulation of recombinant P/Q-type calcium channels by regulators of G protein signalling proteins. *J. Physiol.* 528 Pt 1: 65-77.

1. Fast synaptic transmission is triggered by the activation of presynaptic Ca²⁺ channels which can be inhibited by Gbetagamma subunits via G protein-coupled receptors (GPCR). Regulators of G protein signalling (RGS) proteins are GTPase-accelerating proteins (GAPs), which are responsible for >100-fold increases in the GTPase activity of G proteins and might be involved in the regulation of presynaptic Ca²⁺ channels. In this study we investigated the effects of RGS2 on G protein modulation of recombinant P/Q-type channels expressed in a human embryonic kidney (HEK293) cell line using whole-cell recordings. 2. RGS2 markedly accelerates transmitter-mediated inhibition and recovery from inhibition of Ba²⁺ currents (IBa) through P/Q-type channels heterologously expressed with the muscarinic acetylcholine receptor M2 (mAChR M2). 3. Both RGS2 and RGS4 modulate the prepulse facilitation properties of P/Q-type Ca²⁺ channels. G protein reinhibition is accelerated, while release from inhibition is slowed. These kinetics depend on the availability of G protein alpha and betagamma subunits which is altered by RGS proteins. 4. RGS proteins unmask the Ca²⁺ channel beta subunit modulation of Ca²⁺ channel G protein inhibition. In the presence of RGS2, P/Q-type channels containing the beta2a and beta3 subunits reveal significantly altered

THE UNIVERSITY OF CHICAGO PRESS

Panel 4D Summary: Ag3046 Expression of this gene is low/undetectable (CTs > 35)
5 across all of the samples on this panel (data not shown).

10

15

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gcagctcattcagacctatgag-3'	22	847	1333
Probe	TET-5'-ctctcctgccaccctatgacactg-3'-TAMRA	25	883	1334
Reverse	5'-cgacaacaggtacatcatgaag-3'	22	913	1335

20

Panel 4D Summary: Ag3051 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

20

Table BZA. Probe Name Ag3057

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-aacattggaaggacaggagtct-3'	22	2314	1336
Probe	TET-5'-ccccaggagatgtatcagattcagct-3'-TAMRA	26	2336	1337
Reverse	5'-cagatccccaagaacctta-3'	20	2382	1338

Table BZB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3057, Run 211012795	Tissue Name	Rel. Exp.(%) Ag3057, Run 211012795
AD 1 Hippo	12.9	Control (Path) 3 Temporal Ctx	8.4
AD 2 Hippo	27.2	Control (Path) 4 Temporal Ctx	25.0
AD 3 Hippo	11.2	AD 1 Occipital Ctx	29.5
AD 4 Hippo	13.0	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	23.8	AD 3 Occipital Ctx	15.4
AD 6 Hippo	77.4	AD 4 Occipital Ctx	28.9
Control 2 Hippo	34.9	AD 5 Occipital Ctx	15.9
Control 4 Hippo	15.5	AD 6 Occipital Ctx	14.7
Control (Path) 3 Hippo	11.2	Control 1 Occipital Ctx	8.0
AD 1 Temporal Ctx	45.7	Control 2 Occipital Ctx	28.3
AD 2 Temporal Ctx	31.4	Control 3 Occipital Ctx	14.8
AD 3 Temporal Ctx	19.3	Control 4 Occipital Ctx	23.8
AD 4 Temporal Ctx	28.1	Control (Path) 1 Occipital Ctx	73.7
AD 5 Inf Temporal Ctx	41.2	Control (Path) 2 Occipital Ctx	16.8
AD 5 Sup Temporal Ctx	23.3	Control (Path) 3 Occipital Ctx	7.3
AD 6 Inf Temporal Ctx	100.0	Control (Path) 4 Occipital Ctx	13.0
AD 6 Sup Temporal Ctx	75.3	Control 1 Parietal Ctx	14.8
Control 1 Temporal Ctx	9.7	Control 2 Parietal Ctx	29.3
Control 2 Temporal Ctx	21.8	Control 3 Parietal Ctx	12.8
Control 3 Temporal Ctx	16.0	Control (Path) 1 Parietal Ctx	42.6
Control 4 Temporal Ctx	9.3	Control (Path) 2 Parietal Ctx	30.4
Control (Path) 1 Temporal Ctx	45.4	Control (Path) 3 Parietal Ctx	9.7
Control (Path) 2 Temporal Ctx	28.9	Control (Path) 4 Parietal Ctx	30.8

Table BZC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3057, Run 165519995	Tissue Name	Rel. Exp.(%) Ag3057, Run 165519995
Liver adenocarcinoma	9.9	Kidney (fetal)	6.1
Pancreas	3.8	Renal ca. 786-0	8.8
Pancreatic ca. CAPAN 2	40.9	Renal ca. A498	15.1
Adrenal gland	14.5	Renal ca. RXF 393	13.0
Thyroid	6.2	Renal ca. ACHN	2.1
Salivary gland	10.8	Renal ca. UO-31	13.1
Pituitary gland	10.2	Renal ca. TK-10	3.9
Brain (fetal)	51.4	Liver	1.9
Brain (whole)	100.0	Liver (fetal)	4.8
Brain (amygdala)	48.0	Liver ca. (hepatoblast) HepG2	17.3
Brain (cerebellum)	49.3	Lung	10.4
Brain (hippocampus)	47.6	Lung (fetal)	7.2
Brain (substantia nigra)	70.2	Lung ca. (small cell) LX-1	7.6
Brain (thalamus)	51.8	Lung ca. (small cell) NCI-H69	1.0
Cerebral Cortex	11.3	Lung ca. (s.cell var.) SHP-77	7.3
Spinal cord	85.9	Lung ca. (large cell)NCI-H460	16.6
glio/astro U87-MG	16.5	Lung ca. (non-sm. cell) A549	4.5
glio/astro U-118-MG	27.9	Lung ca. (non-s.cell) NCI-H23	6.1
astrocytoma SW1783	13.5	Lung ca. (non-s.cell) HOP-62	8.5
neuro*; met SK-N-AS	12.2	Lung ca. (non-s.cl) NCI-H522	3.0
astrocytoma SF-539	14.2	Lung ca. (squam.) SW 900	8.3
astrocytoma SNB-75	20.9	Lung ca. (squam.) NCI-H596	5.9
glioma SNB-19	23.0	Mammary gland	17.3
glioma U251	25.7	Breast ca.* (pl.ef) MCF-7	7.3
glioma SF-295	9.8	Breast ca.* (pl.ef) MDA-MB-231	35.8
Heart (fetal)	0.2	Breast ca.* (pl.ef) T47D	5.5
Heart	11.1	Breast ca. BT-549	32.1

Skeletal muscle (fetal)	3.2	Breast ca. MDA-N	3.3
Skeletal muscle	18.9	Ovary	1.9
Bone marrow	3.8	Ovarian ca. OVCAR-3	3.8
Thymus	1.5	Ovarian ca. OVCAR-4	6.9
Spleen	5.6	Ovarian ca. OVCAR-5	7.7
Lymph node	7.3	Ovarian ca. OVCAR-8	5.7
Colorectal	9.2	Ovarian ca. IGROV-1	2.5
Stomach	15.0	Ovarian ca.* (ascites) SK-OV-3	20.9
Small intestine	10.4	Uterus	23.5
Colon ca. SW480	5.6	Placenta	6.0
Colon ca.* SW620(SW480 met)	4.9	Prostate	3.4
Colon ca. HT29	2.8	Prostate ca.* (bone met)PC-3	8.4
Colon ca. HCT-116	3.7	Testis	9.8
Colon ca. CaCo-2	10.3	Melanoma Hs688(A).T	6.1
Colon ca. tissue(ODO3866)	9.2	Melanoma* (met) Hs688(B).T	6.0
Colon ca. HCC-2998	5.3	Melanoma UACC-62	1.8
Gastric ca.* (liver met) NCI-N87	34.9	Melanoma M14	14.5
Bladder	14.0	Melanoma LOX IMVI	0.9
Trachea	6.5	Melanoma* (met) SK-MEL-5	3.8
Kidney	3.0	Adipose	12.8

Table BZD. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3057, Run 163577596	Tissue Name	Rel. Exp.(%) Ag3057, Run 163577596
Normal Colon	81.8	Kidney Margin 8120608	1.1
CC Well to Mod Diff (ODO3866)	10.7	Kidney Cancer 8120613	8.1
CC Margin (ODO3866)	17.6	Kidney Margin 8120614	1.1

CC Gr.2 rectosigmoid (ODO3868)	11.2	Kidney Cancer 9010320	2.7
CC Margin (ODO3868)	4.8	Kidney Margin 9010321	2.8
CC Mod Diff (ODO3920)	15.5	Normal Uterus	10.7
CC Margin (ODO3920)	20.2	Uterus Cancer 064011	31.2
CC Gr.2 ascend colon (ODO3921)	36.9	Normal Thyroid	19.1
CC Margin (ODO3921)	9.7	Thyroid Cancer 064010	7.6
CC from Partial Hepatectomy (ODO4309) Mets	26.2	Thyroid Cancer A302152	5.6
Liver Margin (ODO4309)	13.1	Thyroid Margin A302153	13.7
Colon mets to lung (OD04451-01)	2.9	Normal Breast	41.5
Lung Margin (OD04451-02)	6.0	Breast Cancer (OD04566)	7.5
Normal Prostate 6546-1	60.3	Breast Cancer (OD04590-01)	40.6
Prostate Cancer (OD04410)	20.2	Breast Cancer Mets (OD04590-03)	32.3
Prostate Margin (OD04410)	24.5	Breast Cancer Metastasis (OD04655-05)	17.9
Prostate Cancer (OD04720-01)	23.8	Breast Cancer 064006	15.7
Prostate Margin (OD04720-02)	37.4	Breast Cancer 1024	10.2
Normal Lung 061010	42.6	Breast Cancer 9100266	6.5
Lung Met to Muscle (ODO4286)	38.7	Breast Margin 9100265	6.4
Muscle Margin (ODO4286)	9.2	Breast Cancer A209073	20.2
Lung Malignant Cancer (OD03126)	20.7	Breast Margin A2090734	15.2
Lung Margin (OD03126)	17.8	Normal Liver	7.4
Lung Cancer (OD04404)	36.9	Liver Cancer 064003	7.2
Lung Margin (OD04404)	11.0	Liver Cancer 1025	3.2
Lung Cancer (OD04565)	11.3	Liver Cancer 1026	3.5
Lung Margin (OD04565)	10.6	Liver Cancer 6004-T	4.8
Lung Cancer (OD04237-01)	30.4	Liver Tissue 6004-N	4.8

Lung Margin (OD04237-02)	21.0	Liver Cancer 6005-T	2.5
Ocular Mel Met to Liver (ODO4310)	11.7	Liver Tissue 6005-N	1.1
Liver Margin (ODO4310)	13.5	Normal Bladder	32.1
Melanoma Mets to Lung (OD04321)	10.2	Bladder Cancer 1023	5.6
Lung Margin (OD04321)	36.3	Bladder Cancer A302173	49.3
Normal Kidney	100.0	Bladder Cancer (OD04718-01)	28.7
Kidney Ca, Nuclear grade 2 (OD04338)	40.6	Bladder Normal Adjacent (OD04718-03)	24.8
Kidney Margin (OD04338)	16.8	Normal Ovary	6.5
Kidney Ca Nuclear grade 1/2 (OD04339)	11.0	Ovarian Cancer 064008	21.5
Kidney Margin (OD04339)	21.9	Ovarian Cancer (OD04768-07)	35.8
Kidney Ca, Clear cell type (OD04340)	66.4	Ovary Margin (OD04768-08)	5.4
Kidney Margin (OD04340)	18.0	Normal Stomach	80.1
Kidney Ca, Nuclear grade 3 (OD04348)	3.4	Gastric Cancer 9060358	2.2
Kidney Margin (OD04348)	10.2	Stomach Margin 9060359	12.3
Kidney Cancer (OD04622-01)	12.3	Gastric Cancer 9060395	23.2
Kidney Margin (OD04622-03)	1.2	Stomach Margin 9060394	14.6
Kidney Cancer (OD04450-01)	26.1	Gastric Cancer 9060397	11.8
Kidney Margin (OD04450-03)	21.5	Stomach Margin 9060396	5.8
Kidney Cancer 8120607	2.2	Gastric Cancer 064005	55.5

Table BZE. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag3057, Run 171694175	Tissue Name	Rel. Exp.(%) Ag3057, Run 171694175
BA4 Control	7.9	BA17 PSP	21.3
BA4 Control2	16.0	BA17 PSP2	5.4
BA4	2.9	Sub Nigra Control	58.6

Alzheimer's2			
BA4 Parkinson's	32.8	Sub Nigra Control2	25.9
BA4 Parkinson's2	46.7	Sub Nigra Alzheimer's2	20.0
BA4 Huntington's	25.9	Sub Nigra Parkinson's2	85.3
BA4 Huntington's2	4.1	Sub Nigra Huntington's	100.0
BA4 PSP	5.7	Sub Nigra Huntington's2	59.0
BA4 PSP2	36.9	Sub Nigra PSP2	25.2
BA4 Depression	15.6	Sub Nigra Depression	35.8
BA4 Depression2	27.2	Sub Nigra Depression2	26.1
BA7 Control	12.9	Glob Palladus Control	54.3
BA7 Control2	12.4	Glob Palladus Control2	15.2
BA7 Alzheimer's2	5.2	Glob Palladus Alzheimer's	17.4
BA7 Parkinson's	15.4	Glob Palladus Alzheimer's2	15.6
BA7 Parkinson's2	23.2	Glob Palladus Parkinson's	73.7
BA7 Huntington's	34.9	Glob Palladus Parkinson's2	15.8
BA7 Huntington's2	56.3	Glob Palladus PSP	17.7
BA7 PSP	29.1	Glob Palladus PSP2	8.4
BA7 PSP2	12.6	Glob Palladus Depression	24.1
BA7 Depression	12.2	Temp Pole Control	0.8
BA9 Control	4.4	Temp Pole Control2	13.9
BA9 Control2	36.1	Temp Pole Alzheimer's	2.7
BA9 Alzheimer's	3.8	Temp Pole Alzheimer's2	2.0
BA9 Alzheimer's2	2.1	Temp Pole Parkinson's	15.9
BA9 Parkinson's	17.2	Temp Pole Parkinson's2	12.9
BA9 Parkinson's2	20.7	Temp Pole Huntington's	19.2
BA9	36.9	Temp Pole PSP	6.2

Huntington's			
BA9 Huntington's2	15.2	Temp Pole PSP2	1.8
BA9 PSP	27.4	Temp Pole Depression2	10.8
BA9 PSP2	14.0	Cing Gyr Control	35.6
BA9 Depression	7.5	Cing Gyr Control2	15.2
BA9 Depression2	11.7	Cing Gyr Alzheimer's	15.7
BA17 Control	20.7	Cing Gyr Alzheimer's2	10.7
BA17 Control2	15.3	Cing Gyr Parkinson's	47.0
BA17 Alzheimer's2	4.2	Cing Gyr Parkinson's2	48.6
BA17 Parkinson's	41.8	Cing Gyr Huntington's	85.3
BA17 Parkinson's2	27.5	Cing Gyr Huntington's2	53.2
BA17 Huntington's	17.3	Cing Gyr PSP	80.1
BA17 Huntington's2	27.0	Cing Gyr PSP2	6.9
BA17 Depression	81.8	Cing Gyr Depression	18.6
BA17 Depression2	51.4	Cing Gyr Depression2	69.7

CNS_neurodegeneration_v1.0 Summary: Ag3057 The NOV85 gene is found to be slightly but significantly ($p=0.016$) upregulated in the Alzheimer's disease (AD) temporal cortex. The temporal cortex is the region of the brain where neurons degenerate in the mid stages of AD. This increase in expression is not apparent in the occipital cortex, which does not experience neurodegeneration in AD. Since the upregulation of this gene appears to be neurodegeneration-specific both within an individual brain and between brains, this gene is an excellent small molecule target. Therefore, treatment with an antagonist may decrease the pathology seen in Alzheimer's disease.

Panel 1.3D Summary: Ag3057 Highest expression of the NOV85 gene is seen in the CNS. Please see CNS_Neurodegeneration for discussion of utility of this gene in the central nervous system.

Among tissues with metabolic function, this gene has low levels of expression in pancreas, adrenal, thyroid, pituitary, skeletal muscle and adipose. Therefore, modulation of

this gene product may be a treatment for metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

In addition, this gene is expressed at low levels in the cancer cell lines in this panel. This difference in expression is particularly prominent in the CNS cancer cell lines when compared to the normal brain tissues. Thus, this gene could potentially be used as a diagnostic marker in CNS cancers.

Panel 2D Summary: Ag3057 The NOV85 gene is expressed at moderate to low levels in this panel. A higher level of expression is observed in lung, kidney, uterine, gastric and ovarian cancer when compared to the normal adjacent lung, kidney, uterine, gastric and ovarian tissues in this panel. Thus, this gene could be used as a diagnostic marker of cancer in these tissues. Furthermore, inhibition of the activity of this gene product using small molecule drugs may be useful for the treatment of cancer in these tissues

Panel 4D Summary: Ag3057 The amp plot indicates that there is a high probability of experimental failure. (Data not shown.)

Panel CNS_1 Summary: Ag3057 These results confirm expression of the NOV85 gene in the brain. Please see CNS_Neurodegeneration for discussion of utility of this gene in the central nervous system.

NOV86: GTPASE-ACTIVATING PROTEIN

Expression of gene NOV86 was assessed using the primer-probe set Ag3058, described in Table CAA. Results of the RTQ-PCR runs are shown in Tables CAB, CAC and CAD.

Table CAA. Probe Name Ag3058

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agtacccgctgctgaacac-3'	19	534	1339
Probe	TET-5'-accctcattgccaaaggtcaaagcct-3'-TAMRA	25	578	1340
Reverse	5'-tcattggtgctctcataatgga-3'	22	603	1341

Table CAB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3058, Run 165533238	Tissue Name	Rel. Exp.(%) Ag3058, Run 165533238
Liver adenocarcinoma	5.5	Kidney (fetal)	0.9
Pancreas	1.2	Renal ca. 786-0	0.6
Pancreatic ca. CAPAN 2	3.9	Renal ca. A498	0.1

Adrenal gland	1.8	Renal ca. RXF 393	1.0
Thyroid	1.7	Renal ca. ACHN	0.4
Salivary gland	6.4	Renal ca. UO-31	15.7
Pituitary gland	1.1	Renal ca. TK-10	0.1
Brain (fetal)	0.6	Liver	3.5
Brain (whole)	2.7	Liver (fetal)	4.9
Brain (amygdala)	3.0	Liver ca. (hepatoblast) HepG2	5.3
Brain (cerebellum)	0.4	Lung	16.5
Brain (hippocampus)	2.4	Lung (fetal)	8.4
Brain (substantia nigra)	4.7	Lung ca. (small cell) LX-1	1.8
Brain (thalamus)	2.7	Lung ca. (small cell) NCI-H69	1.0
Cerebral Cortex	1.2	Lung ca. (s.cell var.) SHP-77	1.3
Spinal cord	4.2	Lung ca. (large cell)NCI-H460	0.9
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.1
glio/astro U-118-MG	0.2	Lung ca. (non-s.cell) NCI-H23	0.2
astrocytoma SW1783	0.1	Lung ca. (non-s.cell) HOP-62	4.6
neuro*; met SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
astrocytoma SF-539	2.5	Lung ca. (squam.) SW 900	6.9
astrocytoma SNB-75	5.4	Lung ca. (squam.) NCI-H596	0.5
glioma SNB-19	0.7	Mammary gland	1.7
glioma U251	1.1	Breast ca.* (pl.ef) MCF-7	0.0
glioma SF-295	0.3	Breast ca.* (pl.ef) MDA-MB-231	19.2
Heart (fetal)	1.0	Breast ca.* (pl.ef) T47D	2.4
Heart	1.4	Breast ca. BT-549	4.7
Skeletal muscle (fetal)	0.5	Breast ca. MDA-N	0.9
Skeletal muscle	1.0	Ovary	1.8
Bone marrow	32.1	Ovarian ca. OVCAR-3	0.3
Thymus	29.3	Ovarian ca. OVCAR-4	0.3
Spleen	46.3	Ovarian ca. OVCAR-	9.0

		5	
Lymph node	100.0	Ovarian ca. OVCAR-8	0.7
Colorectal	1.4	Ovarian ca. IGROV-1	0.7
Stomach	12.1	Ovarian ca.* (ascites) SK-OV-3	0.2
Small intestine	16.6	Uterus	2.4
Colon ca. SW480	2.3	Placenta	13.8
Colon ca.* SW620(SW480 met)	0.9	Prostate	0.9
Colon ca. HT29	1.2	Prostate ca.* (bone met)PC-3	3.1
Colon ca. HCT-116	2.9	Testis	0.2
Colon ca. CaCo-2	3.5	Melanoma Hs688(A).T	0.1
Colon ca. tissue(ODO3866)	4.6	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	2.8	Melanoma UACC-62	0.1
Gastric ca.* (liver met) NCI-N87	4.6	Melanoma M14	7.2
Bladder	1.9	Melanoma LOX IMVI	1.6
Trachea	8.1	Melanoma* (met) SK-MEL-5	0.1
Kidney	0.3	Adipose	3.5

Table CAC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3058, Run 162569974	Tissue Name	Rel. Exp.(%) Ag3058, Run 162569974
Normal Colon	30.8	Kidney Margin 8120608	6.5
CC Well to Mod Diff (ODO3866)	6.8	Kidney Cancer 8120613	2.9
CC Margin (ODO3866)	6.7	Kidney Margin 8120614	6.3
CC Gr.2 rectosigmoid (ODO3868)	8.5	Kidney Cancer 9010320	35.4
CC Margin (ODO3868)	1.3	Kidney Margin 9010321	17.0
CC Mod Diff (ODO3920)	9.9	Normal Uterus	2.4
CC Margin (ODO3920)	11.3	Uterus Cancer 064011	10.2
CC Gr.2 ascend colon	14.7	Normal Thyroid	6.8

(ODO3921)			
CC Margin (ODO3921)	10.7	Thyroid Cancer 064010	5.1
CC from Partial Hepatectomy (ODO4309) Mets	18.7	Thyroid Cancer A302152	6.8
Liver Margin (ODO4309)	12.4	Thyroid Margin A302153	7.9
Colon mets to lung (OD04451-01)	12.7	Normal Breast	17.7
Lung Margin (OD04451- 02)	18.7	Breast Cancer (OD04566)	10.4
Normal Prostate 6546-1	13.8	Breast Cancer (OD04590-01)	18.2
Prostate Cancer (OD04410)	13.8	Breast Cancer Mets (OD04590-03)	59.9
Prostate Margin (OD04410)	11.7	Breast Cancer Metastasis (OD04655-05)	52.5
Prostate Cancer (OD04720-01)	7.3	Breast Cancer 064006	21.3
Prostate Margin (OD04720-02)	19.2	Breast Cancer 1024	9.7
Normal Lung 061010	100.0	Breast Cancer 9100266	12.9
Lung Met to Muscle (ODO4286)	29.5	Breast Margin 9100265	14.1
Muscle Margin (ODO4286)	8.2	Breast Cancer A209073	11.6
Lung Malignant Cancer (OD03126)	31.6	Breast Margin A2090734	6.6
Lung Margin (OD03126)	48.6	Normal Liver	7.2
Lung Cancer (OD04404)	16.4	Liver Cancer 064003	5.4
Lung Margin (OD04404)	17.1	Liver Cancer 1025	6.8
Lung Cancer (OD04565)	6.7	Liver Cancer 1026	13.2
Lung Margin (OD04565)	20.3	Liver Cancer 6004-T	7.2
Lung Cancer (OD04237- 01)	24.0	Liver Tissue 6004-N	6.7
Lung Margin (OD04237- 02)	43.2	Liver Cancer 6005-T	14.7
Ocular Mel Met to Liver (ODO4310)	3.4	Liver Tissue 6005-N	4.5
Liver Margin (ODO4310)	8.2	Normal Bladder	20.2
Melanoma Mets to Lung (OD04321)	6.0	Bladder Cancer 1023	9.2

		TNFalpha + IL-1beta	
Primary Th2 act	3.5	Microvascular Dermal EC none	38.7
Primary Tr1 act	8.3	Microvascular Dermal EC TNFalpha + IL-1beta	24.7
Primary Th1 rest	7.6	Bronchial epithelium TNFalpha + IL1beta	9.2
Primary Th2 rest	3.1	Small airway epithelium none	13.8
Primary Tr1 rest	4.8	Small airway epithelium TNFalpha + IL-1beta	100.0
CD45RA CD4 lymphocyte act	13.7	Coronary artery SMC rest	22.2
CD45RO CD4 lymphocyte act	7.3	Coronary artery SMC TNFalpha + IL-1beta	11.3
CD8 lymphocyte act	2.3	Astrocytes rest	14.7
Secondary CD8 lymphocyte rest	3.4	Astrocytes TNFalpha + IL-1beta	12.0
Secondary CD8 lymphocyte act	4.4	KU-812 (Basophil) rest	12.2
CD4 lymphocyte none	1.5	KU-812 (Basophil) PMA/ionomycin	46.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.3	CCD1106 (Keratinocytes) none	24.0
LAK cells rest	15.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.6
LAK cells IL-2	4.5	Liver cirrhosis	3.7
LAK cells IL-2+IL-12	4.5	Lupus kidney	3.9
LAK cells IL-2+IFN gamma	8.1	NCI-H292 none	46.3
LAK cells IL-2+ IL-18	7.9	NCI-H292 IL-4	57.4
LAK cells PMA/ionomycin	10.2	NCI-H292 IL-9	63.3
NK Cells IL-2 rest	2.0	NCI-H292 IL-13	28.1
Two Way MLR 3 day	2.6	NCI-H292 IFN gamma	25.9
Two Way MLR 5 day	3.1	HPAEC none	14.8
Two Way MLR 7 day	2.3	HPAEC TNF alpha + IL-1 beta	19.9
PBMC rest	2.4	Lung fibroblast none	29.1
PBMC PWM	12.3	Lung fibroblast TNF alpha + IL-1 beta	13.3
PBMC PHA-L	5.1	Lung fibroblast IL-4	39.8
Ramos (B cell) none	0.0	Lung fibroblast IL-9	24.5
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	19.1

B lymphocytes PWM	32.3	Lung fibroblast IFN gamma	45.7
B lymphocytes CD40L and IL-4	4.1	Dermal fibroblast CCD1070 rest	57.4
EOL-1 dbcAMP	0.7	Dermal fibroblast CCD1070 TNF alpha	81.8
EOL-1 dbcAMP PMA/ionomycin	6.7	Dermal fibroblast CCD1070 IL-1 beta	25.0
Dendritic cells none	12.8	Dermal fibroblast IFN gamma	51.1
Dendritic cells LPS	11.8	Dermal fibroblast IL-4	52.9
Dendritic cells anti-CD40	13.8	IBD Colitis 2	0.7
Monocytes rest	6.2	IBD Crohn's	3.4
Monocytes LPS	6.5	Colon	22.5
Macrophages rest	18.3	Lung	19.2
Macrophages LPS	10.0	Thymus	24.7
HUVEC none	33.7	Kidney	13.0
HUVEC starved	49.0		

Panel 1.3D Summary: Ag3058 Highest expression of the NOV86 gene, a GTPase-activating protein homolog, is seen in the lymph node (CT=27.8). Among tissues with metabolic function, this gene has low levels of expression in pancreas, adrenal, thyroid, pituitary, heart, skeletal muscle, liver and adipose. Rab GTPases are integral to vesicular transport in the secretory and endocytic pathways. Therefore, therapeutic modulation of this gene product may be a treatment for metabolic and endocrine diseases, including obesity and Types 1 and 2 diabetes.

This GTPase activating enzyme like molecule is also expressed at low levels in the CNS. Thus, it may be useful in treating diseases of the nervous system, stroke or CNS trauma.

In addition, this gene is expressed at low levels in the cancer cell lines in this panel. Therefore, modulation of expression of this gene may be useful in treating cancer.

Panel 2D Summary: Ag3058 The NOV86 gene is expressed at low levels in this panel. There is higher expression in kidney, breast, liver and bladder cancer samples compared to the adjacent normal tissue. Conversely, there is lower expression in lung cancer samples compared to the adjacent normal tissue. Thus, the expression of this gene could be used as a diagnostic marker for kidney, breast, liver, bladder and lung cancers. Furthermore, modulation of expression of this gene may also be used for therapy of these cancers.

Panel 4D Summary: Ag3058 The NOV86 gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response and tissue response

in health and disease, with the highest expression being detected in TNF alpha plus IL-1 beta treated small airway epithelial cells (CT=28.03). Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as COPD, emphysema, asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis.

NOV87a and NOV87b

Expression of gene NOV87a and full length clone NOV87b was assessed using the primer-probe set Ag3040, described in Table CBA.

Table CBA. Probe Name Ag3040

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gagccctgaagctcttctttc-3'	21	617	1342
Probe	TET-5'-cttctcgcaacttcgccagttcatt-3'-TAMRA	25	663	1343
Reverse	5'-cctggctcctgctcactgat-3'	19	694	1344

CNS_neurodegeneration_v1.0 Summary: Ag3040 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag3040 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag3040 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

NOV88

Expression of gene NOV88 was assessed using the primer-probe set Ag2923, described in Table CCA.

Table CCA. Probe Name Ag2923

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-agtcaacagatttgccacat-3'	21	45	1345
Probe	TET-5'-tcaccagggctgcttttaactctggt-3'-TAMRA	26	77	1346
Reverse	5'-cagtgaaggggtcactgatg-3'	20	120	1347

CNS_neurodegeneration_v1.0 Summary: Ag2923 Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag2923 Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

5 **Panel 2D Summary:** Ag2923 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag2923 Expression of this gene is low/undetectable (CTs > 34.5) across all of the samples on this panel (data not shown).

NOV89

10 Expression of gene NOV89 was assessed using the primer-probe set Ag2924, described in Table CDA.

Table CDA. Probe Name Ag2924

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gaacgggaagcttggttatcaat-3'	22	231	1348
Probe	TET-5'-agatctcaccaaaatcaaaggggca-3'-TAMRA	26	282	1349
Reverse	5'-atgatgtactcagtgccagcat-3'	22	308	1350

15 **CNS_neurodegeneration_v1.0 Summary:** Ag2924 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 1.3D Summary: Ag2924 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

Panel 4D Summary: Ag2924 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

20 **Panel 5 Islet Summary:** Ag2924 Run 242285280 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown). Run 243564308 The amp plot indicates that there were experimental difficulties with this run.

NOV90

25 Expression of gene NOV90 was assessed using the primer-probe set Ag3045, described in Table CEA. Results of the RTQ-PCR runs are shown in Tables CEB, CEC, CED, CEE, CEF and CEG.

Table CEA. Probe Name Ag3045

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-caccggatcatatgaaatcaat-3'	22	605	1351
Probe	TET-5'-tgtaattgacctgttctctgcaccag-3'- TAMRA	26	639	1352
Reverse	5'-ccaccatcaacatttgaatca-3'	21	669	1353

Table CEB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3045, Run 211012233	Tissue Name	Rel. Exp.(%) Ag3045, Run 211012233
AD 1 Hippo	0.8	Control (Path) 3 Temporal Ctx	0.2
AD 2 Hippo	0.1	Control (Path) 4 Temporal Ctx	4.4
AD 3 Hippo	0.5	AD 1 Occipital Ctx	2.1
AD 4 Hippo	0.3	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	8.5	AD 3 Occipital Ctx	0.3
AD 6 Hippo	0.3	AD 4 Occipital Ctx	2.0
Control 2 Hippo	0.5	AD 5 Occipital Ctx	0.8
Control 4 Hippo	0.5	AD 6 Occipital Ctx	1.1
Control (Path) 3 Hippo	0.4	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	1.4	Control 2 Occipital Ctx	2.2
AD 2 Temporal Ctx	0.0	Control 3 Occipital Ctx	2.1
AD 3 Temporal Ctx	0.7	Control 4 Occipital Ctx	0.2
AD 4 Temporal Ctx	2.7	Control (Path) 1 Occipital Ctx	4.1
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	2.4
AD 5 SupTemporal Ctx	4.4	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	0.9	Control (Path) 4 Occipital Ctx	2.7
AD 6 Sup Temporal Ctx	0.9	Control 1 Parietal Ctx	0.0
Control 1 Temporal Ctx	0.0	Control 2 Parietal Ctx	4.1
Control 2 Temporal Ctx	0.7	Control 3 Parietal Ctx	1.5
Control 3 Temporal	1.5	Control (Path) 1	2.6

Ctx		Parietal Ctx	
Control 4 Temporal Ctx	0.6	Control (Path) 2 Parietal Ctx	2.6
Control (Path) 1 Temporal Ctx	3.1	Control (Path) 3 Parietal Ctx	0.2
Control (Path) 2 Temporal Ctx	3.6	Control (Path) 4 Parietal Ctx	3.4

Table CEC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3045, Run 165519994	Tissue Name	Rel. Exp.(%) Ag3045, Run 165519994
Liver adenocarcinoma	10.3	Kidney (fetal)	1.9
Pancreas	3.7	Renal ca. 786-0	4.8
Pancreatic ca. CAPAN 2	3.5	Renal ca. A498	16.2
Adrenal gland	7.5	Renal ca. RXF 393	6.6
Thyroid	2.2	Renal ca. ACHN	1.0
Salivary gland	8.6	Renal ca. UO-31	6.7
Pituitary gland	28.7	Renal ca. TK-10	1.1
Brain (fetal)	20.3	Liver	4.4
Brain (whole)	100.0	Liver (fetal)	1.9
Brain (amygdala)	8.4	Liver ca. (hepatoblast) HepG2	3.6
Brain (cerebellum)	50.7	Lung	4.1
Brain (hippocampus)	36.6	Lung (fetal)	2.3
Brain (substantia nigra)	12.1	Lung ca. (small cell) LX-1	0.8
Brain (thalamus)	29.7	Lung ca. (small cell) NCI-H69	2.5
Cerebral Cortex	4.2	Lung ca. (s.cell var.) SHP-77	0.6
Spinal cord	10.4	Lung ca. (large cell) NCI-H460	2.1
glio/astro U87-MG	3.2	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	6.6	Lung ca. (non-s.cell) NCI-H23	3.2
astrocytoma SW1783	2.4	Lung ca. (non-s.cell) HOP-62	4.5
neuro*; met SK-N-AS	3.5	Lung ca. (non-s.cl) NCI-H522	0.6
astrocytoma SF-539	35.4	Lung ca. (squam.) SW 900	16.4
astrocytoma SNB-75	14.0	Lung ca. (squam.)	2.6

		NCI-H596	
glioma SNB-19	37.9	Mammary gland	2.9
glioma U251	44.4	Breast ca.* (pl.ef) MCF-7	4.8
glioma SF-295	0.4	Breast ca.* (pl.ef) MDA-MB-231	7.1
Heart (fetal)	0.4	Breast ca.* (pl.ef) T47D	8.5
Heart	9.5	Breast ca. BT-549	3.8
Skeletal muscle (fetal)	1.0	Breast ca. MDA-N	1.3
Skeletal muscle	5.4	Ovary	1.3
Bone marrow	9.9	Ovarian ca. OVCAR-3	2.8
Thymus	8.8	Ovarian ca. OVCAR-4	0.9
Spleen	9.4	Ovarian ca. OVCAR-5	3.6
Lymph node	21.8	Ovarian ca. OVCAR-8	20.2
Colorectal	3.3	Ovarian ca. IGROV-1	0.6
Stomach	6.3	Ovarian ca.* (ascites) SK-OV-3	11.4
Small intestine	8.0	Uterus	18.8
Colon ca. SW480	4.1	Placenta	7.0
Colon ca.* SW620(SW480 met)	2.1	Prostate	15.7
Colon ca. HT29	0.5	Prostate ca.* (bone met)PC-3	2.8
Colon ca. HCT-116	2.5	Testis	11.4
Colon ca. CaCo-2	1.9	Melanoma Hs688(A).T	0.4
Colon ca. tissue(ODO3866)	2.4	Melanoma* (met) Hs688(B).T	0.4
Colon ca. HCC-2998	2.4	Melanoma UACC-62	3.1
Gastric ca.* (liver met) NCI-N87	3.3	Melanoma M14	7.9
Bladder	11.3	Melanoma LOX IMVI	0.0
Trachea	1.7	Melanoma* (met) SK-MEL-5	1.8
Kidney	10.2	Adipose	1.4

Table CED. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3045, Run 163577595	Tissue Name	Rel. Exp.(%) Ag3045, Run 163577595
Normal Colon	30.4	Kidney Margin 8120608	1.1
CC Well to Mod Diff (ODO3866)	2.9	Kidney Cancer 8120613	2.5
CC Margin (ODO3866)	1.1	Kidney Margin 8120614	0.2
CC Gr.2 rectosigmoid (ODO3868)	2.7	Kidney Cancer 9010320	2.9
CC Margin (ODO3868)	4.1	Kidney Margin 9010321	4.0
CC Mod Diff (ODO3920)	1.9	Normal Uterus	2.1
CC Margin (ODO3920)	3.9	Uterus Cancer 064011	22.2
CC Gr.2 ascend colon (ODO3921)	2.7	Normal Thyroid	7.1
CC Margin (ODO3921)	2.6	Thyroid Cancer 064010	2.0
CC from Partial Hepatectomy (ODO4309) Mets	3.6	Thyroid Cancer A302152	2.0
Liver Margin (ODO4309)	3.6	Thyroid Margin A302153	20.9
Colon mets to lung (OD04451-01)	1.8	Normal Breast	9.6
Lung Margin (OD04451- 02)	2.5	Breast Cancer (OD04566)	5.3
Normal Prostate 6546-1	100.0	Breast Cancer (OD04590-01)	55.5
Prostate Cancer (OD04410)	39.0	Breast Cancer Mets (OD04590-03)	92.7
Prostate Margin (OD04410)	17.3	Breast Cancer Metastasis (OD04655-05)	31.9
Prostate Cancer (OD04720-01)	15.1	Breast Cancer 064006	21.8
Prostate Margin (OD04720-02)	37.1	Breast Cancer 1024	6.5
Normal Lung 061010	47.0	Breast Cancer 9100266	8.4
Lung Met to Muscle (ODO4286)	4.4	Breast Margin 9100265	4.0
Muscle Margin (ODO4286)	10.3	Breast Cancer A209073	12.2

Lung Malignant Cancer (OD03126)	8.1	Breast Margin A2090734	2.2
Lung Margin (OD03126)	8.8	Normal Liver	1.0
Lung Cancer (OD04404)	47.0	Liver Cancer 064003	0.8
Lung Margin (OD04404)	11.0	Liver Cancer 1025	0.7
Lung Cancer (OD04565)	6.9	Liver Cancer 1026	0.6
Lung Margin (OD04565)	1.7	Liver Cancer 6004-T	3.1
Lung Cancer (OD04237-01)	12.2	Liver Tissue 6004-N	1.7
Lung Margin (OD04237-02)	5.4	Liver Cancer 6005-T	0.8
Ocular Mel Met to Liver (ODO4310)	5.1	Liver Tissue 6005-N	0.8
Liver Margin (ODO4310)	1.0	Normal Bladder	12.5
Melanoma Mets to Lung (OD04321)	3.3	Bladder Cancer 1023	1.4
Lung Margin (OD04321)	5.7	Bladder Cancer A302173	6.7
Normal Kidney	7.2	Bladder Cancer (OD04718-01)	16.4
Kidney Ca, Nuclear grade 2 (OD04338)	8.4	Bladder Normal Adjacent (OD04718-03)	23.3
Kidney Margin (OD04338)	2.9	Normal Ovary	2.0
Kidney Ca Nuclear grade 1/2 (OD04339)	0.8	Ovarian Cancer 064008	5.6
Kidney Margin (OD04339)	1.8	Ovarian Cancer (OD04768-07)	2.4
Kidney Ca, Clear cell type (OD04340)	2.1	Ovary Margin (OD04768-08)	2.4
Kidney Margin (OD04340)	4.2	Normal Stomach	1.0
Kidney Ca, Nuclear grade 3 (OD04348)	1.6	Gastric Cancer 9060358	0.5
Kidney Margin (OD04348)	10.4	Stomach Margin 9060359	2.4
Kidney Cancer (OD04622-01)	1.3	Gastric Cancer 9060395	4.5
Kidney Margin (OD04622-03)	1.1	Stomach Margin 9060394	1.9
Kidney Cancer (OD04450-01)	9.0	Gastric Cancer 9060397	7.7
Kidney Margin (OD04450-03)	14.0	Stomach Margin 9060396	0.3

Kidney Cancer 8120607	0.9	Gastric Cancer 064005	8.3
-----------------------	-----	--------------------------	-----

Table CEE. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3045, Run 164886427	Tissue Name	Rel. Exp.(%) Ag3045, Run 164886427
Daoy- Medulloblastoma	3.4	Ca Ski- Cervical epidermoid carcinoma (metastasis)	6.6
TE671- Medulloblastoma	4.4	ES-2- Ovarian clear cell carcinoma	2.1
D283 Med- Medulloblastoma	4.2	Ramos- Stimulated with PMA/ionomycin 6h	6.0
PFSK-1- Primitive Neuroectodermal	1.7	Ramos- Stimulated with PMA/ionomycin 14h	1.7
XF-498- CNS	7.1	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	4.2
SNB-78- Glioma	2.2	Raji- Burkitt's lymphoma	2.0
SF-268- Glioblastoma	0.8	Daudi- Burkitt's lymphoma	3.4
T98G- Glioblastoma	1.3	U266- B-cell plasmacytoma	4.9
SK-N-SH- Neuroblastoma (metastasis)	3.1	CA46- Burkitt's lymphoma	0.5
SF-295- Glioblastoma	2.0	RL- non-Hodgkin's B-cell lymphoma	2.1
Cerebellum	100.0	JM1- pre-B-cell lymphoma	5.8
Cerebellum	1.4	Jurkat- T cell leukemia	3.5
NCI-H292- Mucoepidermoid lung carcinoma	23.0	TF-1- Erythroleukemia	8.3
DMS-114- Small cell lung cancer	8.4	HUT 78- T-cell lymphoma	3.4
DMS-79- Small cell lung cancer	26.6	U937- Histiocytic lymphoma	1.2
NCI-H146- Small cell lung cancer	5.0	KU-812- Myelogenous leukemia	2.6
NCI-H526- Small cell lung cancer	30.6	769-P- Clear cell renal carcinoma	1.4
NCI-N417- Small cell lung cancer	1.3	Caki-2- Clear cell renal carcinoma	3.3
NCI-H82- Small cell lung cancer	4.2	SW 839- Clear cell renal carcinoma	1.6
NCI-H157- Squamous cell lung cancer	2.6	G401- Wilms' tumor	2.0

OVCAR-5- Ovarian carcinoma	1.4	SCC-9- Squamous cell carcinoma of tongue	0.3
RL95-2- Uterine carcinoma	0.3	SCC-15- Squamous cell carcinoma of tongue	0.7
HelaS3- Cervical adenocarcinoma	2.3	CAL 27- Squamous cell carcinoma of tongue	8.0

Table CEF. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3045, Run 162559632	Tissue Name	Rel. Exp.(%) Ag3045, Run 162559632
Secondary Th1 act	10.0	HUVEC IL-1beta	2.2
Secondary Th2 act	12.0	HUVEC IFN gamma	4.2
Secondary Tr1 act	21.9	HUVEC TNF alpha + IFN gamma	3.2
Secondary Th1 rest	3.3	HUVEC TNF alpha + IL4	4.4
Secondary Th2 rest	5.6	HUVEC IL-11	2.9
Secondary Tr1 rest	6.7	Lung Microvascular EC none	6.9
Primary Th1 act	13.1	Lung Microvascular EC TNFalpha + IL-1beta	4.2
Primary Th2 act	16.2	Microvascular Dermal EC none	8.2
Primary Tr1 act	17.0	Microvascular Dermal EC TNFalpha + IL-1beta	4.8
Primary Th1 rest	47.0	Bronchial epithelium TNFalpha + IL1beta	1.1
Primary Th2 rest	21.9	Small airway epithelium none	2.3
Primary Tr1 rest	33.2	Small airway epithelium TNFalpha + IL-1beta	34.2
CD45RA CD4 lymphocyte act	5.1	Coronary artery SMC rest	2.4
CD45RO CD4 lymphocyte act	11.6	Coronary artery SMC TNFalpha + IL-1beta	2.1
CD8 lymphocyte act	8.6	Astrocytes rest	5.4
Secondary CD8 lymphocyte rest	8.1	Astrocytes TNFalpha + IL-1beta	3.6
Secondary CD8 lymphocyte act	6.0	KU-812 (Basophil) rest	4.3
CD4 lymphocyte none	2.3	KU-812 (Basophil) PMA/ionomycin	14.1
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.0	CCD1106 (Keratinocytes) none	6.0

LAK cells rest	10.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.5
LAK cells IL-2	8.6	Liver cirrhosis	3.3
LAK cells IL-2+IL-12	8.5	Lupus kidney	1.7
LAK cells IL-2+IFN gamma	14.7	NCI-H292 none	100.0
LAK cells IL-2+ IL-18	18.0	NCI-H292 IL-4	84.1
LAK cells PMA/ionomycin	3.7	NCI-H292 IL-9	96.6
NK Cells IL-2 rest	6.8	NCI-H292 IL-13	24.1
Two Way MLR 3 day	6.6	NCI-H292 IFN gamma	22.8
Two Way MLR 5 day	8.8	HPAEC none	4.9
Two Way MLR 7 day	5.7	HPAEC TNF alpha + IL-1 beta	3.9
PBMC rest	4.5	Lung fibroblast none	3.8
PBMC PWM	27.5	Lung fibroblast TNF alpha + IL-1 beta	2.6
PBMC PHA-L	7.0	Lung fibroblast IL-4	9.5
Ramos (B cell) none	6.6	Lung fibroblast IL-9	5.5
Ramos (B cell) ionomycin	34.4	Lung fibroblast IL-13	6.0
B lymphocytes PWM	29.1	Lung fibroblast IFN gamma	5.2
B lymphocytes CD40L and IL-4	10.7	Dermal fibroblast CCD1070 rest	11.4
EOL-1 dbcAMP	6.3	Dermal fibroblast CCD1070 TNF alpha	32.5
EOL-1 dbcAMP PMA/ionomycin	12.2	Dermal fibroblast CCD1070 IL-1 beta	5.7
Dendritic cells none	9.7	Dermal fibroblast IFN gamma	3.6
Dendritic cells LPS	6.1	Dermal fibroblast IL-4	5.9
Dendritic cells anti- CD40	11.7	IBD Colitis 2	0.3
Monocytes rest	10.2	IBD Crohn's	0.4
Monocytes LPS	9.5	Colon	8.7
Macrophages rest	22.5	Lung	8.0
Macrophages LPS	2.4	Thymus	13.5
HUVEC none	5.7	Kidney	43.5
HUVEC starved	9.1		

Table CEG. Panel CNS_1

Tissue Name	Rel. Exp.(%) Ag3045,	Tissue Name	Rel. Exp.(%) Ag3045,
-------------	----------------------	-------------	----------------------

	Run 171694539		Run 171694539
BA4 Control	0.9	BA17 PSP	0.0
BA4 Control2	0.0	BA17 PSP2	6.5
BA4 Alzheimer's2	9.3	Sub Nigra Control	2.6
BA4 Parkinson's	55.1	Sub Nigra Control2	0.0
BA4 Parkinson's2	19.1	Sub Nigra Alzheimer's2	11.2
BA4 Huntington's	18.7	Sub Nigra Parkinson's2	9.0
BA4 Huntington's2	6.5	Sub Nigra Huntington's	22.8
BA4 PSP	0.0	Sub Nigra Huntington's2	21.9
BA4 PSP2	6.1	Sub Nigra PSP2	2.3
BA4 Depression	15.1	Sub Nigra Depression	0.0
BA4 Depression2	6.6	Sub Nigra Depression2	3.0
BA7 Control	8.7	Glob Palladus Control	2.1
BA7 Control2	0.0	Glob Palladus Control2	0.0
BA7 Alzheimer's2	14.7	Glob Palladus Alzheimer's	0.9
BA7 Parkinson's	26.8	Glob Palladus Alzheimer's2	2.8
BA7 Parkinson's2	16.3	Glob Palladus Parkinson's	61.6
BA7 Huntington's	61.1	Glob Palladus Parkinson's2	2.9
BA7 Huntington's2	23.5	Glob Palladus PSP	0.0
BA7 PSP	0.0	Glob Palladus PSP2	0.8
BA7 PSP2	11.6	Glob Palladus Depression	0.0
BA7 Depression	8.1	Temp Pole Control	0.0
BA9 Control	0.9	Temp Pole Control2	1.3
BA9 Control2	0.0	Temp Pole Alzheimer's	0.0
BA9 Alzheimer's	0.0	Temp Pole Alzheimer's2	5.6
BA9 Alzheimer's2	20.6	Temp Pole Parkinson's	27.9
BA9 Parkinson's	31.4	Temp Pole	2.0

		Parkinson's2	
BA9 Parkinson's2	21.9	Temp Pole Huntington's	29.1
BA9 Huntington's	23.0	Temp Pole PSP	0.0
BA9 Huntington's2	15.2	Temp Pole PSP2	0.9
BA9 PSP	0.0	Temp Pole Depression2	5.5
BA9 PSP2	3.6	Cing Gyr Control	2.0
BA9 Depression	2.0	Cing Gyr Control2	0.9
BA9 Depression2	2.9	Cing Gyr Alzheimer's	0.0
BA17 Control	15.1	Cing Gyr Alzheimer's2	14.5
BA17 Control2	3.3	Cing Gyr Parkinson's	26.4
BA17 Alzheimer's2	15.5	Cing Gyr Parkinson's2	12.0
BA17 Parkinson's	100.0	Cing Gyr Huntington's	46.0
BA17 Parkinson's2	54.0	Cing Gyr Huntington's2	19.8
BA17 Huntington's	32.3	Cing Gyr PSP	0.5
BA17 Huntington's2	5.9	Cing Gyr PSP2	1.9
BA17 Depression	4.0	Cing Gyr Depression	3.9
BA17 Depression2	5.6	Cing Gyr Depression2	8.4

CNS_neurodegeneration_v1.0 Summary: Ag3045 The NOV90 gene is not differentially expressed in the postmortem brains of Alzheimer's diseased patients when compared to non-demented control. However, this panel does confirm the expression of this gene in the CNS of an independent sample of individuals. See panel 1 for a discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3045 Expression of the NOV90 gene shows a brain-preferential expression profile, and is expressed at moderate levels in all regions examined. Thus, this gene may be of utility as a small molecule target in neurologic disease.

In addition, significant expression is seen in a cluster of brain cancer cell lines. Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker to detect the presence of brain cancer. Furthermore, therapeutic

modulation of the expression or function of this gene may be effective in the treatment of brain cancer.

Panel 2D Summary: Ag3045 The NOV90 gene is expressed at low levels in this panel. There is higher expression in gastric, breast, uterus and lung cancers than the normal samples from these organs. Expression of this gene could therefore be used as a diagnostic marker for gastric, lung, breast and uterine cancers. Furthermore, modulation of the gene product using small molecule inhibitors could be used for the treatment of these cancers.

Panel 3D Summary: Ag3045 The NOV90 gene is expressed at a low level in most of the cancer cell lines on this panel. Modulation of the gene product using small molecule inhibitors could therefore be used for the treatment of cancer. Highest expression of this gene is seen in the cerebellum, confirming the results seen in Panel 1.3D.

Panel 4D Summary: Ag3045 The NOV90 gene, a serine/threonine-protein kinase homolog is expressed at moderate levels in pulmonary mucoepidermoid cells prepared under several conditions of cell activation: NCI-H292 none (CT=29.01), NCI-H292 IL-4 (CT=29.26), NCI-H292 IL-9 (CT=29.06), NCI-H292 IL-13 (CT=31.06), NCI-H292 IFN gamma (CT=31.14). Therefore, small molecule antagonists that block the function of the NOV90 gene product may be useful in several autoimmune and inflammatory diseases of the lung including, but not limited to, chronic obstructive pulmonary disease, asthma, and emphysema.

Panel CNS_1 Summary: Ag3045 The expression in this panel confirms the presence of the NOV90 in the brain. Thus, this gene may be of utility as a small molecule target in neurologic disease.

NOV91

Expression of gene NOV91 was assessed using the primer-probe set Ag3018, described in Table CFA. Results of the RTQ-PCR runs are shown in Tables CFB, CFC and CFD.

Table CFA. Probe Name Ag3018

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ctgcaggggttgagaaatg-3'	19	71	1354
Probe	TET-5'-ccatcctgggcaaaccgaaggat-3'-TAMRA	23	107	1355
Reverse	5'-ctacacccatcatgttcacatg-3'	22	130	1356

Table CFB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3018, Run 209820676	Tissue Name	Rel. Exp.(%) Ag3018, Run 209820676
AD 1 Hippo	0.0	Control (Path) 3 Temporal Ctx	4.5
AD 2 Hippo	12.5	Control (Path) 4 Temporal Ctx	22.7
AD 3 Hippo	2.4	AD 1 Occipital Ctx	11.3
AD 4 Hippo	7.7	AD 2 Occipital Ctx (Missing)	42.6
AD 5 hippo	72.7	AD 3 Occipital Ctx	3.0
AD 6 Hippo	47.3	AD 4 Occipital Ctx	14.2
Control 2 Hippo	12.9	AD 5 Occipital Ctx	52.1
Control 4 Hippo	3.6	AD 6 Occipital Ctx	48.0
Control (Path) 3 Hippo	3.8	Control 1 Occipital Ctx	1.6
AD 1 Temporal Ctx	13.6	Control 2 Occipital Ctx	32.1
AD 2 Temporal Ctx	29.9	Control 3 Occipital Ctx	6.7
AD 3 Temporal Ctx	4.0	Control 4 Occipital Ctx	0.0
AD 4 Temporal Ctx	26.8	Control (Path) 1 Occipital Ctx	73.7
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.6
AD 5 Sup Temporal Ctx	54.7	Control (Path) 3 Occipital Ctx	1.7
AD 6 Inf Temporal Ctx	48.6	Control (Path) 4 Occipital Ctx	23.3
AD 6 Sup Temporal Ctx	49.0	Control 1 Parietal Ctx	6.5
Control 1 Temporal Ctx	1.7	Control 2 Parietal Ctx	57.8
Control 2 Temporal Ctx	9.9	Control 3 Parietal Ctx	7.3
Control 3 Temporal Ctx	14.4	Control (Path) 1 Parietal Ctx	57.8
Control 4 Temporal Ctx	2.3	Control (Path) 2 Parietal Ctx	21.2
Control (Path) 1 Temporal Ctx	30.4	Control (Path) 3 Parietal Ctx	3.6
Control (Path) 2 Temporal Ctx	36.9	Control (Path) 4 Parietal Ctx	17.6

Table CFC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3018, Run 167819112	Tissue Name	Rel. Exp.(%) Ag3018, Run 167819112
Liver adenocarcinoma	13.0	Kidney (fetal)	7.3
Pancreas	6.0	Renal ca. 786-0	9.9
Pancreatic ca. CAPAN 2	12.9	Renal ca. A498	4.7
Adrenal gland	1.3	Renal ca. RXF 393	21.2
Thyroid	1.3	Renal ca. ACHN	9.5
Salivary gland	0.0	Renal ca. UO-31	4.1
Pituitary gland	7.9	Renal ca. TK-10	12.7
Brain (fetal)	7.9	Liver	2.6
Brain (whole)	44.1	Liver (fetal)	2.9
Brain (amygdala)	3.2	Liver ca. (hepatoblast) HepG2	10.2
Brain (cerebellum)	65.1	Lung	6.7
Brain (hippocampus)	9.8	Lung (fetal)	14.9
Brain (substantia nigra)	4.9	Lung ca. (small cell) LX-1	26.4
Brain (thalamus)	8.8	Lung ca. (small cell) NCI-H69	4.6
Cerebral Cortex	10.7	Lung ca. (s.cell var.) SHP-77	75.3
Spinal cord	7.6	Lung ca. (large cell)NCI-H460	2.9
glio/astro U87-MG	23.2	Lung ca. (non-sm. cell) A549	55.1
glio/astro U-118-MG	37.9	Lung ca. (non-s.cell) NCI-H23	9.6
astrocytoma SW1783	35.1	Lung ca. (non-s.cell) HOP-62	8.6
neuro*; met SK-N-AS	2.7	Lung ca. (non-s.cl) NCI-H522	11.0
astrocytoma SF-539	21.2	Lung ca. (squam.) SW 900	4.7
astrocytoma SNB-75	10.6	Lung ca. (squam.) NCI-H596	17.6
glioma SNB-19	24.8	Mammary gland	2.2
glioma U251	28.5	Breast ca.* (pl.ef) MCF-7	17.3
glioma SF-295	34.6	Breast ca.* (pl.ef) MDA-MB-231	17.3
Heart (fetal)	6.4	Breast ca.* (pl.ef) T47D	6.5
Heart	16.2	Breast ca. BT-549	6.3

Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	8.4
Skeletal muscle	11.7	Ovary	0.0
Bone marrow	9.2	Ovarian ca. OVCAR-3	6.2
Thymus	27.7	Ovarian ca. OVCAR-4	9.7
Spleen	5.6	Ovarian ca. OVCAR-5	25.2
Lymph node	5.8	Ovarian ca. OVCAR-8	20.3
Colorectal	15.0	Ovarian ca. IGROV-1	9.2
Stomach	1.3	Ovarian ca.* (ascites) SK-OV-3	26.6
Small intestine	1.7	Uterus	1.4
Colon ca. SW480	14.7	Placenta	0.0
Colon ca.* SW620(SW480 met)	100.0	Prostate	4.2
Colon ca. HT29	2.1	Prostate ca.* (bone met)PC-3	15.3
Colon ca. HCT-116	12.4	Testis	4.5
Colon ca. CaCo-2	22.7	Melanoma Hs688(A).T	3.9
Colon ca. tissue(ODO3866)	7.0	Melanoma* (met) Hs688(B).T	9.8
Colon ca. HCC-2998	35.6	Melanoma UACC-62	4.1
Gastric ca.* (liver met) NCI-N87	19.6	Melanoma M14	7.9
Bladder	20.3	Melanoma LOX IMVI	11.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.7
Kidney	5.7	Adipose	0.0

Table CFD. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3018, Run 164528110	Tissue Name	Rel. Exp.(%) Ag3018, Run 164528110
Secondary Th1 act	14.9	HUVEC IL-1beta	7.9
Secondary Th2 act	22.1	HUVEC IFN gamma	4.9
Secondary Tr1 act	28.9	HUVEC TNF alpha + IFN gamma	4.5
Secondary Th1 rest	4.1	HUVEC TNF alpha + IL4	6.7
Secondary Th2 rest	5.9	HUVEC IL-11	4.7

Secondary Tr1 rest	4.4	Lung Microvascular EC none	10.9
Primary Th1 act	19.5	Lung Microvascular EC TNFalpha + IL-1beta	10.1
Primary Th2 act	9.4	Microvascular Dermal EC none	13.2
Primary Tr1 act	13.6	Microvascular Dermal EC TNFalpha + IL-1beta	6.7
Primary Th1 rest	45.7	Bronchial epithelium TNFalpha + IL1beta	6.2
Primary Th2 rest	18.6	Small airway epithelium none	2.4
Primary Tr1 rest	16.5	Small airway epithelium TNFalpha + IL-1beta	17.6
CD45RA CD4 lymphocyte act	5.6	Coronary artery SMC rest	3.5
CD45RO CD4 lymphocyte act	17.1	Coronary artery SMC TNFalpha + IL-1beta	1.0
CD8 lymphocyte act	16.5	Astrocytes rest	7.3
Secondary CD8 lymphocyte rest	9.6	Astrocytes TNFalpha + IL-1beta	3.1
Secondary CD8 lymphocyte act	5.3	KU-812 (Basophil) rest	2.0
CD4 lymphocyte none	4.6	KU-812 (Basophil) PMA/ionomycin	6.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	12.0	CCD1106 (Keratinocytes) none	4.1
LAK cells rest	13.5	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.8
LAK cells IL-2	18.6	Liver cirrhosis	2.1
LAK cells IL-2+IL-12	14.6	Lupus kidney	1.7
LAK cells IL-2+IFN gamma	16.3	NCI-H292 none	13.5
LAK cells IL-2+ IL-18	33.9	NCI-H292 IL-4	14.5
LAK cells PMA/ionomycin	4.5	NCI-H292 IL-9	23.3
NK Cells IL-2 rest	17.3	NCI-H292 IL-13	8.0
Two Way MLR 3 day	15.4	NCI-H292 IFN gamma	6.4
Two Way MLR 5 day	6.4	HPAEC none	4.6
Two Way MLR 7 day	6.3	HPAEC TNF alpha + IL-1 beta	6.4
PBMC rest	4.2	Lung fibroblast none	5.0
PBMC PWM	51.1	Lung fibroblast TNF alpha + IL-1 beta	1.5
PBMC PHA-L	14.7	Lung fibroblast IL-4	13.6

Ramos (B cell) none	24.3	Lung fibroblast IL-9	7.2
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	13.1
B lymphocytes PWM	34.4	Lung fibroblast IFN gamma	6.8
B lymphocytes CD40L and IL-4	33.0	Dermal fibroblast CCD1070 rest	15.4
EOL-1 dbcAMP	13.3	Dermal fibroblast CCD1070 TNF alpha	44.8
EOL-1 dbcAMP PMA/ionomycin	7.4	Dermal fibroblast CCD1070 IL-1 beta	5.9
Dendritic cells none	3.0	Dermal fibroblast IFN gamma	3.6
Dendritic cells LPS	8.8	Dermal fibroblast IL-4	9.7
Dendritic cells anti-CD40	6.0	IBD Colitis 2	0.9
Monocytes rest	5.2	IBD Crohn's	0.0
Monocytes LPS	2.6	Colon	6.3
Macrophages rest	4.2	Lung	3.7
Macrophages LPS	1.3	Thymus	9.9
HUVEC none	5.3	Kidney	27.2
HUVEC starved	12.8		

CNS_neurodegeneration_v1.0 Summary: Ag3018 This panel does not show differential expression of the NOV91 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

Panel 1.3D Summary: Ag3018 The NOV91 gene is expressed in the brain at low levels. D-dopachrome tautomerase has been implicated in the production of neuromelanin from the toxic quinone products of dopamine, and this pathway has been implicated in the death of dopaminergic neurons in Parkinson's disease. Thus, this gene may represent an excellent small molecule target for the treatment or prevention of Parkinson's disease.

In addition, significant expression is seen in a cluster of lung, brain, and colon cancer cell lines. Thus, expression of this gene could be used to differentiate between these sample and other samples on this panel and as a marker to detect the presence of these cancers.

Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of lung, brain, and colon cancer.

References:

Matsunaga J, Sinha D, Solano F, Santis C, Wistow G, Hearing V. Macrophage migration inhibitory factor (MIF)--its role in catecholamine metabolism. *Cell Mol Biol (Noisy-le-grand)* 1999 Nov;45(7):1035-40

Macrophage migration inhibitory factor (MIF) was originally identified several
 5 decades ago as a lymphokine-derived protein that inhibited monocyte migration. Recently, it
 has been reported that MIF has D-dopachrome tautomerase, phenylpyruvate tautomerase and
 thiol protein oxidoreductase activities, although the physiological significance of those
 activities is not yet clear. Here we show that MIF is able to catalyze the conversion of
 dopaminechrome and norepinephrinechrome, toxic quinone products of the neurotransmitters
 10 dopamine and norepinephrine, respectively, to indole derivatives that may serve as precursors
 to neuromelanin. Since MIF is highly expressed in human brain, these observations raise the
 possibility that MIF participates in a detoxification pathway for catecholamine products and
 could therefore have an important role for neural tissues. The potential role of MIF in the
 formation of neuromelanin from catecholamines is also an extremely interesting possibility.

15 Drukarch B, van Muiswinkel FL. Neuroprotection for Parkinson's disease: a new
 approach for a new millennium. *Expert Opin Investig Drugs* 2001 Oct;10(10):1855-68

Parkinson's disease (PD) is the only neurodegenerative disorder in which
 pharmacological intervention has resulted in a marked decrease in morbidity and a significant
 delay in mortality. However, the medium to long-term efficacy of this pharmacotherapy,
 20 mainly consisting of dopaminomimetics like L -dopa and dopamine receptor agonists, suffers
 greatly from the unrelenting progression of the disease process underlying PD, i.e., the
 degeneration of neuromelanin-containing, dopaminergic neurones in the substantia nigra.
 Efforts concentrated on understanding the mechanisms of dopaminergic cell death in
 Parkinson's disease have led to identification of a large variety of pathogenetic factors,
 25 including excessive release of oxygen free radicals during enzymatic dopamine breakdown,
 impairment of mitochondrial function, production of inflammatory mediators, loss of trophic
 support, and apoptosis. Therapeutic approaches aimed at correcting these abnormalities are
 currently being evaluated on their efficacy as neuroprotectants for PD. Here, we focus on the
 process of dopamine auto-oxidation, the chain of reactions leading to the formation of
 30 neuromelanin, as an often overlooked, yet obvious pathogenetic factor. In particular, we
 discuss the option of drug-mediated stimulation of endogenous mechanisms responsible for
 the detoxification of dopamine auto-oxidation products as a novel means of neuroprotection in
 Parkinson's disease.

Panel 4D Summary: Ag3018 The NOV91 gene, a D-dopachrome tautomerase homolog, is widely expressed in this panel, with highest expression in Ramos (B cells) activated by treatment with ionomycin (CT=31.28). Therefore, small molecule antagonists that block the function of the NOV91 gene product may be useful in several autoimmune and inflammatory diseases in which activated B cells can play major roles as sources of autoantibody-producing cells and also as powerful antigen-presenting cells, including, but not limited to, Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

Panel 5 Islet Summary: Ag3018 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

NOV92

Expression of gene NOV92 was assessed using the primer-probe set Ag3048, described in Table CGA. Results of the RTQ-PCR runs are shown in Tables CGB, CGC, CGD and CGE.

Table CGA. Probe Name Ag3048

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-cacaaccagctgacagacagt-3'	21	490	1357
Probe	TET-5'-ccaccaccttcagcaagctgcatag-3'-TAMRA	25	521	1358
Reverse	5'-gggagagatccaggtattcaag-3'	22	547	1359

Table CGB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3048, Run 168017062	Tissue Name	Rel. Exp.(%) Ag3048, Run 168017062
Liver adenocarcinoma	2.0	Kidney (fetal)	0.6
Pancreas	0.0	Renal ca. 786-0	3.8
Pancreatic ca. CAPAN 2	0.3	Renal ca. A498	17.7
Adrenal gland	0.0	Renal ca. RXF 393	12.5
Thyroid	0.5	Renal ca. ACHN	0.1
Salivary gland	0.6	Renal ca. UO-31	1.1
Pituitary gland	0.5	Renal ca. TK-10	0.5
Brain (fetal)	0.9	Liver	0.0
Brain (whole)	2.4	Liver (fetal)	0.0
Brain (amygdala)	2.9	Liver ca. (hepatoblast) HepG2	0.2
Brain (cerebellum)	7.0	Lung	1.4

Brain (hippocampus)	2.1	Lung (fetal)	7.1
Brain (substantia nigra)	4.5	Lung ca. (small cell) LX-1	0.2
Brain (thalamus)	4.6	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	1.0	Lung ca. (s.cell var.) SHP-77	0.9
Spinal cord	7.6	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	2.8	Lung ca. (non-sm. cell) A549	0.0
glio/astro U-118-MG	5.9	Lung ca. (non-s.cell) NCI-H23	0.4
astrocytoma SW1783	2.7	Lung ca. (non-s.cell) HOP-62	41.2
neuro*; met SK-N-AS	1.1	Lung ca. (non-s.cl) NCI-H522	1.6
astrocytoma SF-539	100.0	Lung ca. (squam.) SW 900	0.0
astrocytoma SNB-75	41.2	Lung ca. (squam.) NCI-H596	0.0
glioma SNB-19	0.4	Mammary gland	15.9
glioma U251	6.7	Breast ca.* (pl.ef) MCF-7	3.5
glioma SF-295	7.3	Breast ca.* (pl.ef) MDA-MB-231	2.0
Heart (fetal)	5.1	Breast ca.* (pl.ef) T47D	0.3
Heart	1.3	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	4.7	Breast ca. MDA-N	0.3
Skeletal muscle	2.6	Ovary	3.8
Bone marrow	0.9	Ovarian ca. OVCAR- 3	0.2
Thymus	0.9	Ovarian ca. OVCAR- 4	0.2
Spleen	0.3	Ovarian ca. OVCAR- 5	4.3
Lymph node	0.0	Ovarian ca. OVCAR- 8	0.8
Colorectal	2.9	Ovarian ca. IGROV- 1	0.0
Stomach	0.8	Ovarian ca.* (ascites) SK-OV-3	1.5
Small intestine	1.0	Uterus	0.6
Colon ca. SW480	0.8	Placenta	0.9

Colon ca.* SW620(SW480 met)	1.9	Prostate	3.0
Colon ca. HT29	0.3	Prostate ca.* (bone met)PC-3	1.7
Colon ca. HCT-116	0.0	Testis	3.9
Colon ca. CaCo-2	0.2	Melanoma Hs688(A).T	32.8
Colon ca. tissue(ODO3866)	13.2	Melanoma* (met) Hs688(B).T	31.4
Colon ca. HCC-2998	4.1	Melanoma UACC-62	1.8
Gastric ca.* (liver met) NCI-N87	1.5	Melanoma M14	0.0
Bladder	3.0	Melanoma LOX IMVI	0.0
Trachea	3.7	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	4.3

Table CGC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag3048, Run 170858352	Tissue Name	Rel. Exp.(%) Ag3048, Run 170858352
Normal Colon	4.7	Kidney Margin 8120608	0.5
CC Well to Mod Diff (ODO3866)	12.0	Kidney Cancer 8120613	0.1
CC Margin (ODO3866)	1.5	Kidney Margin 8120614	0.4
CC Gr.2 rectosigmoid (ODO3868)	10.1	Kidney Cancer 9010320	13.1
CC Margin (ODO3868)	0.8	Kidney Margin 9010321	0.9
CC Mod Diff (ODO3920)	8.4	Normal Uterus	0.5
CC Margin (ODO3920)	2.6	Uterus Cancer 064011	1.0
CC Gr.2 ascend colon (ODO3921)	13.6	Normal Thyroid	3.3
CC Margin (ODO3921)	2.1	Thyroid Cancer 064010	0.7
CC from Partial Hepatectomy (ODO4309) Mets	3.1	Thyroid Cancer A302152	4.5
Liver Margin (ODO4309)	0.1	Thyroid Margin A302153	0.8
Colon mets to lung (OD04451-01)	3.3	Normal Breast	20.0

Lung Margin (OD04451-02)	1.1	Breast Cancer (OD04566)	6.5
Normal Prostate 6546-1	4.1	Breast Cancer (OD04590-01)	10.4
Prostate Cancer (OD04410)	19.8	Breast Cancer Mets (OD04590-03)	11.4
Prostate Margin (OD04410)	4.5	Breast Cancer Metastasis (OD04655-05)	1.0
Prostate Cancer (OD04720-01)	9.7	Breast Cancer 064006	23.2
Prostate Margin (OD04720-02)	7.5	Breast Cancer 1024	25.7
Normal Lung 061010	3.3	Breast Cancer 9100266	16.4
Lung Met to Muscle (ODO4286)	24.5	Breast Margin 9100265	24.0
Muscle Margin (ODO4286)	1.7	Breast Cancer A209073	32.3
Lung Malignant Cancer (OD03126)	11.6	Breast Margin A2090734	33.2
Lung Margin (OD03126)	3.1	Normal Liver	0.0
Lung Cancer (OD04404)	22.2	Liver Cancer 064003	0.0
Lung Margin (OD04404)	7.2	Liver Cancer 1025	0.0
Lung Cancer (OD04565)	36.1	Liver Cancer 1026	4.2
Lung Margin (OD04565)	2.0	Liver Cancer 6004-T	0.0
Lung Cancer (OD04237-01)	12.6	Liver Tissue 6004-N	2.9
Lung Margin (OD04237-02)	6.7	Liver Cancer 6005-T	2.2
Ocular Mel Met to Liver (ODO4310)	0.0	Liver Tissue 6005-N	0.0
Liver Margin (ODO4310)	0.4	Normal Bladder	23.8
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	11.8
Lung Margin (OD04321)	1.7	Bladder Cancer A302173	7.2
Normal Kidney	0.8	Bladder Cancer (OD04718-01)	12.7
Kidney Ca, Nuclear grade 2 (OD04338)	0.3	Bladder Normal Adjacent (OD04718-03)	4.9
Kidney Margin (OD04338)	1.5	Normal Ovary	3.5
Kidney Ca Nuclear grade 1/2 (OD04339)	0.2	Ovarian Cancer 064008	100.0

Kidney Margin (OD04339)	0.9	Ovarian Cancer (OD04768-07)	2.5
Kidney Ca, Clear cell type (OD04340)	0.4	Ovary Margin (OD04768-08)	8.1
Kidney Margin (OD04340)	1.0	Normal Stomach	1.5
Kidney Ca, Nuclear grade 3 (OD04348)	10.3	Gastric Cancer 9060358	0.5
Kidney Margin (OD04348)	0.4	Stomach Margin 9060359	0.8
Kidney Cancer (OD04622-01)	34.4	Gastric Cancer 9060395	7.5
Kidney Margin (OD04622-03)	0.2	Stomach Margin 9060394	1.9
Kidney Cancer (OD04450-01)	0.0	Gastric Cancer 9060397	27.0
Kidney Margin (OD04450-03)	0.2	Stomach Margin 9060396	1.5
Kidney Cancer 8120607	9.0	Gastric Cancer 064005	3.3

Table CGD. Panel 3D

Tissue Name	Rel. Exp.(%) Ag3048, Run 172133336	Tissue Name	Rel. Exp.(%) Ag3048, Run 172133336
Daoy- Medulloblastoma	6.5	Ca Ski- Cervical epidermoid carcinoma (metastasis)	0.0
TE671- Medulloblastoma	0.0	ES-2- Ovarian clear cell carcinoma	0.0
D283 Med- Medulloblastoma	0.0	Ramos- Stimulated with PMA/ionomycin 6h	0.0
PFSK-1- Primitive Neuroectodermal	18.0	Ramos- Stimulated with PMA/ionomycin 14h	0.0
XF-498- CNS	1.4	MEG-01- Chronic myelogenous leukemia (megokaryoblast)	2.6
SNB-78- Glioma	100.0	Raji- Burkitt's lymphoma	0.0
SF-268- Glioblastoma	0.0	Daudi- Burkitt's lymphoma	0.0
T98G- Glioblastoma	24.1	U266- B-cell plasmacytoma	0.0
SK-N-SH- Neuroblastoma (metastasis)	32.3	CA46- Burkitt's lymphoma	0.0
SF-295- Glioblastoma	2.0	RL- non-Hodgkin's B-cell lymphoma	0.0
Cerebellum	3.0	JM1- pre-B-cell lymphoma	0.0

Cerebellum	4.2	Jurkat- T cell leukemia	0.0
NCI-H292- Mucoepidermoid lung carcinoma	38.4	TF-1- Erythroleukemia	2.3
DMS-114- Small cell lung cancer	1.7	HUT 78- T-cell lymphoma	0.0
DMS-79- Small cell lung cancer	3.0	U937- Histiocytic lymphoma	0.7
NCI-H146- Small cell lung cancer	3.3	KU-812- Myelogenous leukemia	0.0
NCI-H526- Small cell lung cancer	0.0	769-P- Clear cell renal carcinoma	0.0
NCI-N417- Small cell lung cancer	0.0	Caki-2- Clear cell renal carcinoma	2.5
NCI-H82- Small cell lung cancer	0.0	SW 839- Clear cell renal carcinoma	0.0
NCI-H157- Squamous cell lung cancer (metastasis)	1.4	G401- Wilms' tumor	2.3
NCI-H1155- Large cell lung cancer	1.8	Hs766T- Pancreatic carcinoma (LN metastasis)	1.5
NCI-H1299- Large cell lung cancer	2.8	CAPAN-1- Pancreatic adenocarcinoma (liver metastasis)	0.0
NCI-H727- Lung carcinoid	0.0	SU86.86- Pancreatic carcinoma (liver metastasis)	0.8
NCI-UMC-11- Lung carcinoid	0.0	BxPC-3- Pancreatic adenocarcinoma	0.7
LX-1- Small cell lung cancer	0.0	HPAC- Pancreatic adenocarcinoma	0.0
Colo-205- Colon cancer	7.3	MIA PaCa-2- Pancreatic carcinoma	0.7
KM12- Colon cancer	1.6	CFPAC-1- Pancreatic ductal adenocarcinoma	1.7
KM20L2- Colon cancer	0.0	PANC-1- Pancreatic epithelioid ductal carcinoma	6.7
NCI-H716- Colon cancer	0.0	T24- Bladder carcinoma (transitional cell)	6.1
SW-48- Colon adenocarcinoma	0.6	5637- Bladder carcinoma	2.1
SW1116- Colon adenocarcinoma	0.8	HT-1197- Bladder carcinoma	1.7
LS 174T- Colon adenocarcinoma	2.5	UM-UC-3- Bladder carcinoma (transitional cell)	0.7
SW-948- Colon adenocarcinoma	0.0	A204- Rhabdomyosarcoma	5.4

SW-480- Colon adenocarcinoma	2.7	HT-1080- Fibrosarcoma	11.7
NCI-SNU-5- Gastric carcinoma	1.6	MG-63- Osteosarcoma	0.0
KATO III- Gastric carcinoma	1.8	SK-LMS-1- Leiomyosarcoma (vulva)	7.5
NCI-SNU-16- Gastric carcinoma	1.4	SJRH30- Rhabdomyosarcoma (met to bone marrow)	0.0
NCI-SNU-1- Gastric carcinoma	0.0	A431- Epidermoid carcinoma	5.1
RF-1- Gastric adenocarcinoma	0.0	WM266-4- Melanoma	16.6
RF-48- Gastric adenocarcinoma	0.9	DU 145- Prostate carcinoma (brain metastasis)	0.0
MKN-45- Gastric carcinoma	5.6	MDA-MB-468- Breast adenocarcinoma	2.5
NCI-N87- Gastric carcinoma	2.3	SCC-4- Squamous cell carcinoma of tongue	0.0
OVCAR-5- Ovarian carcinoma	4.8	SCC-9- Squamous cell carcinoma of tongue	0.0
RL95-2- Uterine carcinoma	0.0	SCC-15- Squamous cell carcinoma of tongue	0.6
HelaS3- Cervical adenocarcinoma	0.0	CAL 27- Squamous cell carcinoma of tongue	0.5

Table CGE. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3048, Run 164315038	Tissue Name	Rel. Exp.(%) Ag3048, Run 164315038
Secondary Th1 act	0.4	HUVEC IL-1beta	0.0
Secondary Th2 act	0.0	HUVEC IFN gamma	1.9
Secondary Tr1 act	0.6	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	1.2
Secondary Th2 rest	0.0	HUVEC IL-11	0.3
Secondary Tr1 rest	0.0	Lung Microvascular EC none	3.9
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.2
Primary Th2 act	0.0	Microvascular Dermal EC none	0.6
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	1.4
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	28.7

Primary Th2 rest	0.9	Small airway epithelium none	39.2
Primary Tr1 rest	0.6	Small airway epithelium TNFalpha + IL-1beta	84.7
CD45RA CD4 lymphocyte act	8.6	Coronary artery SMC rest	17.9
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	5.5
CD8 lymphocyte act	0.0	Astrocytes rest	6.1
Secondary CD8 lymphocyte rest	0.6	Astrocytes TNFalpha + IL-1beta	5.5
Secondary CD8 lymphocyte act	0.6	KU-812 (Basophil) rest	0.7
CD4 lymphocyte none	0.5	KU-812 (Basophil) PMA/ionomycin	3.8
2ry Th1/Th2/Tr1_anti-CD95 CH11	1.3	CCD1106 (Keratinocytes) none	79.0
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	97.3
LAK cells IL-2	0.0	Liver cirrhosis	4.3
LAK cells IL-2+IL-12	0.0	Lupus kidney	3.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 none	87.1
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-4	77.4
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-9	82.9
NK Cells IL-2 rest	0.0	NCI-H292 IL-13	54.0
Two Way MLR 3 day	0.0	NCI-H292 IFN gamma	49.3
Two Way MLR 5 day	0.0	HPAEC none	0.7
Two Way MLR 7 day	0.8	HPAEC TNF alpha + IL-1 beta	0.0
PBMC rest	0.8	Lung fibroblast none	35.1
PBMC PWM	0.0	Lung fibroblast TNF alpha + IL-1 beta	7.0
PBMC PHA-L	2.2	Lung fibroblast IL-4	46.0
Ramos (B cell) none	0.0	Lung fibroblast IL-9	38.7
Ramos (B cell) ionomycin	0.8	Lung fibroblast IL-13	26.8
B lymphocytes PWM	1.0	Lung fibroblast IFN gamma	68.8
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 rest	100.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	59.9
EOL-1 dbcAMP	0.0	Dermal fibroblast	39.8

PMA/ionomycin		CCD1070 IL-1 beta	
Dendritic cells none	1.8	Dermal fibroblast IFN gamma	17.0
Dendritic cells LPS	1.7	Dermal fibroblast IL-4	30.6
Dendritic cells anti-CD40	6.8	IBD Colitis 2	0.0
Monocytes rest	1.2	IBD Crohn's	0.7
Monocytes LPS	0.0	Colon	2.0
Macrophages rest	1.9	Lung	36.1
Macrophages LPS	1.3	Thymus	0.6
HUVEC none	0.0	Kidney	1.7
HUVEC starved	2.9		

Panel 1.3D Summary: Ag3048 The expression of the NOV92 gene appears to be highest in a sample derived from a brain cancer cell line (SF-539) (CT=29.4). In addition, there is substantial expression associated with samples derived from another brain cancer cell line, two melanoma cell lines and a lung cancer cell line. Thus, the expression of this gene could be used to distinguish SF-539 cells from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be of benefit in treatment of brain or lung cancer or melanoma.

This gene, a leucine-rich repeat homolog, is expressed at low levels in the CNS. The leucine-rich repeat region proteins have been implicated in axonal guidance. Therefore, this gene may be of therapeutic utility in the treatment of any CNS disorder involving neuronal loss, specifically by guiding/enhancing compensatory synaptogenesis and fiber outgrowth, including such clinical conditions as Alzheimer's, Parkinson's, or Huntington's diseases, stroke, head and spinal cord trauma, vascular dementia or spinocerebellar ataxia.

Panel 2D Summary: Ag3048 The expression of the NOV92 gene appears to be highest in a sample derived from an ovarian cancer (CT=29). In addition, there appears to be substantial expression associated with lung cancer, prostate cancer and colon cancer samples. Of note is the differential expression in the lung, colon and prostate cancer samples compared to their respective normal adjacent tissue. Thus, the expression of this gene could be used to distinguish this ovarian cancer sample from other samples in the panel. In addition, the expression of this gene could be used to distinguish colon, prostate or lung cancer samples from their normal adjacent tissue. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be beneficial in the treatment of ovarian, lung, prostate or colon cancer.

Panel 3D Summary: Ag3048 The expression of the NOV92 gene appears to be highest in a sample derived from a brain cancer cell line (SNB-78) (CT=30.2). In addition, there appears to be substantial expression associated with other brain cancer cell line samples and a lung cancer cell line sample. Thus, the expression of this gene could be used to distinguish SNB-78 cells from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be beneficial in the treatment of brain or lung cancer.

Panel 4D Summary: Ag3048 The NOV92 gene, a secreted leucine-rich repeat (LRR) protein, is expressed selectively at moderate levels (CT range 29-31) in several resting and cytokine-activated epithelial and connective tissue cells of lung and skin. Therefore, the NOV92 gene product may be useful as a therapeutic protein as well as a target for therapeutic antibodies and small molecules, each of which may prove beneficial in the reduction or elimination of the symptoms in patients with chronic obstructive pulmonary disease, asthma, emphysema, or psoriasis.

NOV93: IMP DEHYDROGENASE 1

Expression of gene NOV93 was assessed using the primer-probe sets Ag4520 and Ag2904, described in Tables CHA and CHB. Results of the RTQ-PCR runs are shown in Tables CHC, CHD, CHE, CHF, CHG, CHH and CHI.

Table CHA. Probe Name Ag4520

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-ggtagccatgaggatgacaaat-3'	22	758	1360
Probe	TET-5'-acctggacctgctcaccaggttag-3'-TAMRA	24	783	1361
Reverse	5'-cgagtccaagcctatgacatt-3'	21	812	1362

Table CHB. Probe Name Ag2904

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-atacttcaacgacggggataag-3'	22	1300	1363
Probe	TET-5'-ctccatccaggacaaagggtccatt-3'-TAMRA	25	1348	1364
Reverse	5'-aggtagggcacgaacttctg-3'	20	1373	1365

Table CHC. AI_comprehensive panel_v1.0

Tissue Name	Rel. Exp.(%) Ag4520, Run	Tissue Name	Rel. Exp.(%) Ag4520, Run
-------------	-----------------------------	-------------	-----------------------------

	219421380		219421380
110967 COPD-F	11.8	112427 Match Control Psoriasis-F	63.7
110980 COPD-F	12.1	112418 Psoriasis-M	12.5
110968 COPD-M	11.0	112723 Match Control Psoriasis-M	0.3
110977 COPD-M	46.7	112419 Psoriasis-M	10.4
110989 Emphysema-F	34.9	112424 Match Control Psoriasis-M	11.6
110992 Emphysema-F	22.7	112420 Psoriasis-M	52.1
110993 Emphysema-F	21.9	112425 Match Control Psoriasis-M	50.0
110994 Emphysema-F	13.3	104689 (MF) OA Bone-Backus	30.8
110995 Emphysema-F	33.2	104690 (MF) Adj "Normal" Bone-Backus	23.0
110996 Emphysema-F	8.5	104691 (MF) OA Synovium-Backus	22.4
110997 Asthma-M	6.3	104692 (BA) OA Cartilage-Backus	3.4
111001 Asthma-F	32.5	104694 (BA) OA Bone-Backus	17.6
111002 Asthma-F	22.2	104695 (BA) Adj "Normal" Bone-Backus	24.1
111003 Atopic Asthma-F	18.9	104696 (BA) OA Synovium-Backus	13.9
111004 Atopic Asthma-F	20.9	104700 (SS) OA Bone-Backus	21.3
111005 Atopic Asthma-F	10.7	104701 (SS) Adj "Normal" Bone-Backus	21.2
111006 Atopic Asthma-F	3.6	104702 (SS) OA Synovium-Backus	42.3
111417 Allergy-M	13.1	117093 OA Cartilage Rep7	27.9
112347 Allergy-M	2.6	112672 OA Bone5	39.0
112349 Normal Lung-F	2.0	112673 OA Synovium5	14.6
112357 Normal Lung-F	22.4	112674 OA Synovial Fluid cells5	16.6
112354 Normal Lung-M	10.4	117100 OA Cartilage Rep14	5.1
112374 Crohns-F	11.3	112756 OA Bone9	6.8
112389 Match Control Crohns-F	18.6	112757 OA Synovium9	14.7

112375 Crohns-F	8.5	112758 OA Synovial Fluid Cells9	12.8
112732 Match Control Crohns-F	41.8	117125 RA Cartilage Rep2	10.1
112725 Crohns-M	3.2	113492 Bone2 RA	28.7
112387 Match Control Crohns-M	13.5	113493 Synovium2 RA	13.1
112378 Crohns-M	1.6	113494 Syn Fluid Cells RA	26.4
112390 Match Control Crohns-M	49.7	113499 Cartilage4 RA	29.3
112726 Crohns-M	12.8	113500 Bone4 RA	41.2
112731 Match Control Crohns-M	16.7	113501 Synovium4 RA	30.6
112380 Ulcer Col-F	25.5	113502 Syn Fluid Cells4 RA	15.9
112734 Match Control Ulcer Col-F	100.0	113495 Cartilage3 RA	24.1
112384 Ulcer Col-F	48.6	113496 Bone3 RA	31.9
112737 Match Control Ulcer Col-F	7.4	113497 Synovium3 RA	13.7
112386 Ulcer Col-F	4.4	113498 Syn Fluid Cells3 RA	26.4
112738 Match Control Ulcer Col-F	12.9	117106 Normal Cartilage Rep20	6.9
112381 Ulcer Col-M	2.0	113663 Bone3 Normal	2.1
112735 Match Control Ulcer Col-M	12.0	113664 Synovium3 Normal	1.2
112382 Ulcer Col-M	12.1	113665 Syn Fluid Cells3 Normal	0.8
112394 Match Control Ulcer Col-M	4.5	117107 Normal Cartilage Rep22	13.5
112383 Ulcer Col-M	26.8	113667 Bone4 Normal	18.3
112736 Match Control Ulcer Col-M	9.5	113668 Synovium4 Normal	17.6
112423 Psoriasis-F	11.1	113669 Syn Fluid Cells4 Normal	27.4

Table CHD. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag2904, Run 206485416	Rel. Exp.(%) Ag4520, Run 206954220	Tissue Name	Rel. Exp.(%) Ag2904, Run 206485416	Rel. Exp.(%) Ag4520, Run 206954220
AD 1 Hippo	29.3	31.9	Control (Path) 3 Temporal	25.0	28.9

			Ctx		
AD 2 Hippo	40.6	34.2	Control (Path) 4 Temporal Ctx	39.0	30.8
AD 3 Hippo	34.2	26.1	AD 1 Occipital Ctx	27.4	23.7
AD 4 Hippo	29.5	19.1	AD 2 Occipital Ctx (Missing)	0.0	0.0
AD 5 hippo	60.7	58.2	AD 3 Occipital Ctx	33.0	24.8
AD 6 Hippo	80.7	92.7	AD 4 Occipital Ctx	24.5	26.2
Control 2 Hippo	25.0	30.8	AD 5 Occipital Ctx	41.5	30.1
Control 4 Hippo	53.2	40.9	AD 6 Occipital Ctx	30.6	52.1
Control (Path) 3 Hippo	25.0	22.4	Control 1 Occipital Ctx	48.0	60.3
AD 1 Temporal Ctx	52.1	81.2	Control 2 Occipital Ctx	33.4	41.5
AD 2 Temporal Ctx	24.3	21.5	Control 3 Occipital Ctx	25.3	23.3
AD 3 Temporal Ctx	32.3	26.1	Control 4 Occipital Ctx	30.8	25.3
AD 4 Temporal Ctx	40.6	49.7	Control (Path) 1 Occipital Ctx	49.7	37.1
AD 5 Inf Temporal Ctx	52.9	77.4	Control (Path) 2 Occipital Ctx	15.5	12.3
AD 5 SupTemporal Ctx	83.5	100.0	Control (Path) 3 Occipital Ctx	39.2	45.1

AD 6 Inf Temporal Ctx	77.4	85.3	Control (Path) 4 Occipital Ctx	33.4	36.1
AD 6 Sup Temporal Ctx	100.0	87.1	Control 1 Parietal Ctx	37.4	29.1
Control 1 Temporal Ctx	29.9	29.5	Control 2 Parietal Ctx	34.2	52.1
Control 2 Temporal Ctx	20.0	19.1	Control 3 Parietal Ctx	14.0	11.0
Control 3 Temporal Ctx	22.4	14.3	Control (Path) 1 Parietal Ctx	43.2	30.1
Control 4 Temporal Ctx	26.1	10.4	Control (Path) 2 Parietal Ctx	32.8	33.4
Control (Path) 1 Temporal Ctx	46.3	25.7	Control (Path) 3 Parietal Ctx	34.9	36.1
Control (Path) 2 Temporal Ctx	34.2	34.2	Control (Path) 4 Parietal Ctx	48.3	44.4

Table CHE. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4520, Run 219274490	Rel. Exp.(%) Ag4520, Run 219288511	Tissue Name	Rel. Exp.(%) Ag4520, Run 219274490	Rel. Exp.(%) Ag4520, Run 219288511
Adipose	26.6	24.5	Renal ca. TK-10	47.0	42.3
Melanoma* Hs688(A).T	14.1	12.9	Bladder	41.2	37.6
Melanoma* Hs688(B).T	13.6	13.3	Gastric ca. (liver met.) NCI-N87	100.0	100.0
Melanoma* M14	8.5	11.2	Gastric ca. KATO III	14.4	13.5
Melanoma* LOXIMVI	17.2	18.6	Colon ca. SW- 948	5.0	5.3
Melanoma* SK-MEL-5	10.7	11.3	Colon ca. SW480	18.6	20.9
Squamous cell carcinoma SCC-4	4.9	6.6	Colon ca.* (SW480 met) SW620	9.0	9.0
Testis Pool	7.5	7.5	Colon ca. HT29	10.1	10.0
Prostate ca.* (bone met) PC-3	8.0	7.2	Colon ca. HCT- 116	16.7	14.6

Prostate Pool	13.8	13.3	Colon ca. CaCo-2	26.1	26.4
Placenta	18.3	19.5	Colon cancer tissue	21.0	23.8
Uterus Pool	9.0	9.0	Colon ca. SW1116	1.3	2.6
Ovarian ca. OVCAR-3	15.2	13.5	Colon ca. Colo-205	3.5	3.7
Ovarian ca. SK-OV-3	28.9	28.3	Colon ca. SW-48	4.3	4.5
Ovarian ca. OVCAR-4	5.9	6.2	Colon Pool	31.9	30.1
Ovarian ca. OVCAR-5	49.0	49.0	Small Intestine Pool	29.7	32.5
Ovarian ca. IGROV-1	16.3	15.0	Stomach Pool	15.6	15.2
Ovarian ca. OVCAR-8	6.9	6.0	Bone Marrow Pool	13.4	13.9
Ovary	11.3	10.0	Fetal Heart	20.3	21.8
Breast ca. MCF-7	27.7	27.0	Heart Pool	11.3	15.4
Breast ca. MDA-MB-231	37.9	48.3	Lymph Node Pool	26.8	30.1
Breast ca. BT 549	52.9	53.2	Fetal Skeletal Muscle	7.5	10.7
Breast ca. T47D	69.7	74.7	Skeletal Muscle Pool	20.9	24.3
Breast ca. MDA-N	4.7	6.7	Spleen Pool	45.1	57.8
Breast Pool	31.0	37.1	Thymus Pool	31.9	31.0
Trachea	31.0	28.5	CNS cancer (glio/astro) U87-MG	26.1	25.9
Lung	4.5	3.7	CNS cancer (glio/astro) U-118-MG	21.3	24.0
Fetal Lung	68.3	100.0	CNS cancer (neuro;met) SK-N-AS	18.9	21.0
Lung ca. NCI-N417	0.5	0.6	CNS cancer (astro) SF-539	10.6	6.9
Lung ca. LX-1	13.2	11.9	CNS cancer (astro) SNB-75	48.0	47.0
Lung ca. NCI-H146	4.1	3.7	CNS cancer (glio) SNB-19	14.7	13.4
Lung ca.	6.3	6.3	CNS cancer	18.8	20.4

SHP-77			(glio) SF-295		
Lung ca. A549	21.6	23.5	Brain (Amygdala) Pool	2.4	3.3
Lung ca. NCI-H526	1.5	1.7	Brain (cerebellum)	9.2	10.1
Lung ca. NCI-H23	29.7	29.5	Brain (fetal)	7.7	8.2
Lung ca. NCI-H460	14.2	15.1	Brain (Hippocampus) Pool	5.5	4.8
Lung ca. HOP-62	17.4	3.6	Cerebral Cortex Pool	5.4	3.7
Lung ca. NCI-H522	20.3	25.5	Brain (Substantia nigra) Pool	3.8	3.2
Liver	2.0	2.0	Brain (Thalamus) Pool	4.4	4.7
Fetal Liver	14.6	15.9	Brain (whole)	4.0	4.8
Liver ca. HepG2	16.2	18.8	Spinal Cord Pool	6.5	7.0
Kidney Pool	36.3	44.4	Adrenal Gland	28.9	28.3
Fetal Kidney	31.9	29.1	Pituitary gland Pool	3.3	2.5
Renal ca. 786-0	50.7	54.3	Salivary Gland	8.0	6.0
Renal ca. A498	28.9	27.4	Thyroid (female)	9.5	7.6
Renal ca. ACHN	45.4	41.2	Pancreatic ca. CAPAN2	51.4	48.3
Renal ca. UO-31	31.9	31.6	Pancreas Pool	35.4	39.0

Table CHF. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2904, Run 162556421	Tissue Name	Rel. Exp.(%) Ag2904, Run 162556421
Liver adenocarcinoma	16.0	Kidney (fetal)	40.3
Pancreas	6.9	Renal ca. 786-0	13.3
Pancreatic ca. CAPAN 2	6.4	Renal ca. A498	34.2
Adrenal gland	9.7	Renal ca. RXF 393	16.3
Thyroid	8.1	Renal ca. ACHN	21.9
Salivary gland	7.4	Renal ca. UO-31	6.2
Pituitary gland	2.1	Renal ca. TK-10	9.5
Brain (fetal)	1.6	Liver	4.2
Brain (whole)	1.5	Liver (fetal)	11.3

Brain (amygdala)	6.6	Liver ca. (hepatoblast) HepG2	11.3
Brain (cerebellum)	6.7	Lung	70.7
Brain (hippocampus)	7.6	Lung (fetal)	42.3
Brain (substantia nigra)	2.6	Lung ca. (small cell) LX-1	4.7
Brain (thalamus)	5.0	Lung ca. (small cell) NCI-H69	1.6
Cerebral Cortex	10.3	Lung ca. (s.cell var.) SHP-77	2.3
Spinal cord	15.3	Lung ca. (large cell)NCI-H460	6.9
glio/astro U87-MG	13.3	Lung ca. (non-sm. cell) A549	3.5
glio/astro U-118-MG	3.4	Lung ca. (non-s.cell) NCI-H23	10.6
astrocytoma SW1783	6.1	Lung ca. (non-s.cell) HOP-62	4.9
neuro*; met SK-N-AS	6.4	Lung ca. (non-s.cl) NCI-H522	6.9
astrocytoma SF-539	8.0	Lung ca. (squam.) SW 900	18.2
astrocytoma SNB-75	10.9	Lung ca. (squam.) NCI-H596	2.4
glioma SNB-19	7.3	Mammary gland	9.7
glioma U251	3.4	Breast ca.* (pl.ef) MCF-7	18.0
glioma SF-295	8.1	Breast ca.* (pl.ef) MDA-MB-231	10.1
Heart (fetal)	9.9	Breast ca.* (pl.ef) T47D	4.1
Heart	24.3	Breast ca. BT-549	4.8
Skeletal muscle (fetal)	100.0	Breast ca. MDA-N	3.6
Skeletal muscle	9.8	Ovary	18.7
Bone marrow	35.4	Ovarian ca. OVCAR- 3	2.5
Thymus	80.7	Ovarian ca. OVCAR- 4	1.7
Spleen	50.0	Ovarian ca. OVCAR- 5	8.2
Lymph node	29.1	Ovarian ca. OVCAR- 8	3.5
Colorectal	42.3	Ovarian ca. IGROV- 1	4.9
Stomach	10.7	Ovarian ca.* (ascites)	5.2

		SK-OV-3	
Small intestine	25.3	Uterus	10.7
Colon ca. SW480	2.4	Placenta	36.3
Colon ca.* SW620(SW480 met)	3.8	Prostate	10.5
Colon ca. HT29	5.4	Prostate ca.* (bone met)PC-3	1.7
Colon ca. HCT-116	2.6	Testis	7.9
Colon ca. CaCo-2	10.7	Melanoma Hs688(A).T	4.6
Colon ca. tissue(ODO3866)	21.6	Melanoma* (met) Hs688(B).T	6.2
Colon ca. HCC-2998	10.4	Melanoma UACC-62	0.9
Gastric ca.* (liver met) NCI-N87	27.2	Melanoma M14	2.8
Bladder	39.5	Melanoma LOX IMVI	2.9
Trachea	36.9	Melanoma* (met) SK-MEL-5	2.5
Kidney	16.0	Adipose	27.4

Table CHG. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2904, Run 162345750	Tissue Name	Rel. Exp.(%) Ag2904, Run 162345750
Normal Colon	58.6	Kidney Margin 8120608	5.3
CC Well to Mod Diff (ODO3866)	12.2	Kidney Cancer 8120613	7.9
CC Margin (ODO3866)	9.2	Kidney Margin 8120614	14.2
CC Gr.2 rectosigmoid (ODO3868)	50.0	Kidney Cancer 9010320	31.2
CC Margin (ODO3868)	5.6	Kidney Margin 9010321	27.2
CC Mod Diff (ODO3920)	45.4	Normal Uterus	6.5
CC Margin (ODO3920)	27.2	Uterus Cancer 064011	26.6
CC Gr.2 ascend colon (ODO3921)	30.4	Normal Thyroid	10.7
CC Margin (ODO3921)	9.9	Thyroid Cancer 064010	14.2
CC from Partial Hepatectomy (ODO4309) Mets	34.2	Thyroid Cancer A302152	36.3

Liver Margin (ODO4309)	19.1	Thyroid Margin A302153	22.4
Colon mets to lung (OD04451-01)	52.1	Normal Breast	23.5
Lung Margin (OD04451- 02)	43.2	Breast Cancer (OD04566)	6.8
Normal Prostate 6546-1	66.9	Breast Cancer (OD04590-01)	13.5
Prostate Cancer (OD04410)	22.4	Breast Cancer Mets (OD04590-03)	29.3
Prostate Margin (OD04410)	21.9	Breast Cancer Metastasis (OD04655-05)	18.7
Prostate Cancer (OD04720-01)	26.2	Breast Cancer 064006	27.5
Prostate Margin (OD04720-02)	28.9	Breast Cancer 1024	20.9
Normal Lung 061010	100.0	Breast Cancer 9100266	12.2
Lung Met to Muscle (ODO4286)	27.2	Breast Margin 9100265	8.8
Muscle Margin (ODO4286)	13.6	Breast Cancer A209073	15.7
Lung Malignant Cancer (OD03126)	32.5	Breast Margin A2090734	17.6
Lung Margin (OD03126)	63.3	Normal Liver	13.4
Lung Cancer (OD04404)	28.3	Liver Cancer 064003	13.9
Lung Margin (OD04404)	34.9	Liver Cancer 1025	13.9
Lung Cancer (OD04565)	23.3	Liver Cancer 1026	6.6
Lung Margin (OD04565)	49.0	Liver Cancer 6004-T	15.4
Lung Cancer (OD04237- 01)	21.0	Liver Tissue 6004-N	9.6
Lung Margin (OD04237- 02)	55.9	Liver Cancer 6005-T	6.5
Ocular Mel Met to Liver (ODO4310)	3.6	Liver Tissue 6005-N	2.6
Liver Margin (ODO4310)	14.9	Normal Bladder	46.7
Melanoma Mets to Lung (OD04321)	7.0	Bladder Cancer 1023	10.4
Lung Margin (OD04321)	67.8	Bladder Cancer A302173	16.0
Normal Kidney	47.0	Bladder Cancer (OD04718-01)	32.1
Kidney Ca, Nuclear grade 2 (OD04338)	39.5	Bladder Normal Adjacent (OD04718- 03)	28.3

Kidney Margin (OD04338)	45.4	Normal Ovary	2.4
Kidney Ca Nuclear grade 1/2 (OD04339)	80.1	Ovarian Cancer 064008	33.2
Kidney Margin (OD04339)	28.1	Ovarian Cancer (OD04768-07)	80.7
Kidney Ca, Clear cell type (OD04340)	67.8	Ovary Margin (OD04768-08)	11.5
Kidney Margin (OD04340)	57.4	Normal Stomach	27.2
Kidney Ca, Nuclear grade 3 (OD04348)	16.3	Gastric Cancer 9060358	6.9
Kidney Margin (OD04348)	60.7	Stomach Margin 9060359	10.3
Kidney Cancer (OD04622-01)	27.5	Gastric Cancer 9060395	18.4
Kidney Margin (OD04622-03)	5.6	Stomach Margin 9060394	22.8
Kidney Cancer (OD04450-01)	28.5	Gastric Cancer 9060397	27.4
Kidney Margin (OD04450-03)	34.4	Stomach Margin 9060396	9.2
Kidney Cancer 8120607	3.5	Gastric Cancer 064005	77.9

Table CHH. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4520, Run 198263642	Rel. Exp.(%) Ag4520, Run 219310605	Tissue Name	Rel. Exp.(%) Ag4520, Run 198263642	Rel. Exp.(%) Ag4520, Run 219310605
Secondary Th1 act	22.5	26.1	HUVEC IL-1beta	9.5	11.3
Secondary Th2 act	37.4	40.1	HUVEC IFN gamma	17.0	15.6
Secondary Tr1 act	22.8	22.7	HUVEC TNF alpha + IFN gamma	18.0	23.8
Secondary Th1 rest	7.8	6.3	HUVEC TNF alpha + IL4	12.0	12.8
Secondary Th2 rest	8.7	11.7	HUVEC IL-11	5.4	4.6
Secondary Tr1 rest	7.7	3.8	Lung Microvascular EC none	17.6	20.7
Primary Th1 act	35.8	22.4	Lung Microvascular EC	19.9	23.5

			TNFalpha + IL-1beta		
Primary Th2 act	36.3	39.0	Microvascular Dermal EC none	8.1	8.7
Primary Tr1 act	34.9	36.6	Microvascular Dermal EC TNFalpha + IL-1beta	9.1	11.1
Primary Th1 rest	3.8	2.6	Bronchial epithelium TNFalpha + IL1beta	5.3	5.0
Primary Th2 rest	3.4	3.1	Small airway epithelium none	1.6	1.8
Primary Tr1 rest	4.1	8.8	Small airway epithelium TNFalpha + IL-1beta	10.6	9.7
CD45RA CD4 lymphocyte act	15.1	14.0	Coronary artery SMC rest	2.5	1.9
CD45RO CD4 lymphocyte act	24.3	26.2	Coronary artery SMC TNFalpha + IL-1beta	3.7	3.4
CD8 lymphocyte act	13.4	16.3	Astrocytes rest	1.1	2.4
Secondary CD8 lymphocyte rest	16.8	14.1	Astrocytes TNFalpha + IL-1beta	7.1	5.8
Secondary CD8 lymphocyte act	9.1	10.8	KU-812 (Basophil) rest	2.3	1.7
CD4 lymphocyte none	4.2	3.7	KU-812 (Basophil) PMA/ionomycin	8.4	7.0
2ry Th1/Th2/Tr1 anti- CD95 CH11	14.7	13.1	CCD1106 (Keratinocytes) none	1.2	1.7
LAK cells rest	12.4	11.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	6.6	3.6
LAK cells IL-2	14.3	15.5	Liver cirrhosis	2.8	3.6
LAK cells IL-2+IL-12	9.7	6.7	NCI-H292 none	2.6	2.9
LAK cells IL-2+IFN gamma	9.3	9.0	NCI-H292 IL-4	2.6	3.6
LAK cells IL-2+IL-18	12.8	11.7	NCI-H292 IL-9	3.5	3.5

LAK cells PMA/ionomycin	25.0	32.1	NCI-H292 IL-13	3.4	3.1
NK Cells IL-2 rest	18.2	19.6	NCI-H292 IFN gamma	3.6	4.2
Two Way MLR 3 day	20.7	22.7	HPAEC none	5.0	5.2
Two Way MLR 5 day	11.4	15.3	HPAEC TNF alpha + IL-1 beta	19.9	25.0
Two Way MLR 7 day	13.2	14.6	Lung fibroblast none	1.5	1.6
PBMC rest	5.0	4.7	Lung fibroblast TNF alpha + IL-1 beta	3.9	3.2
PBMC PWM	18.0	21.8	Lung fibroblast IL-4	1.2	1.4
PBMC PHA-L	15.0	17.4	Lung fibroblast IL-9	1.1	2.4
Ramos (B cell) none	1.9	2.1	Lung fibroblast IL-13	1.5	1.8
Ramos (B cell) ionomycin	3.0	3.0	Lung fibroblast IFN gamma	2.5	3.1
B lymphocytes PWM	10.4	12.7	Dermal fibroblast CCD1070 rest	3.4	1.8
B lymphocytes CD40L and IL-4	17.0	20.0	Dermal fibroblast CCD1070 TNF alpha	15.3	18.3
EOL-1 dbcAMP	9.8	17.0	Dermal fibroblast CCD1070 IL-1 beta	4.6	4.6
EOL-1 dbcAMP PMA/ionomycin	37.4	32.3	Dermal fibroblast IFN gamma	3.3	2.9
Dendritic cells none	8.4	4.8	Dermal fibroblast IL-4	4.2	2.9
Dendritic cells LPS	20.0	25.2	Dermal Fibroblasts rest	3.3	1.8
Dendritic cells anti- CD40	6.3	6.3	Neutrophils TNFa+LPS	51.8	56.6
Monocytes rest	10.8	9.8	Neutrophils rest	23.2	26.2
Monocytes LPS	100.0	100.0	Colon	1.7	1.6
Macrophages rest	7.6	6.8	Lung	4.0	1.8
Macrophages LPS	29.7	26.6	Thymus	7.9	8.7
HUVEC none	3.1	4.3	Kidney	6.3	3.6
HUVEC starved	10.4	8.0			

Table CHI. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2904, Run 159078059	Tissue Name	Rel. Exp.(%) Ag2904, Run 159078059
Secondary Th1 act	52.5	HUVEC IL-1beta	14.8
Secondary Th2 act	65.5	HUVEC IFN gamma	32.5
Secondary Tr1 act	58.2	HUVEC TNF alpha + IFN gamma	52.9
Secondary Th1 rest	13.1	HUVEC TNF alpha + IL4	28.9
Secondary Th2 rest	18.8	HUVEC IL-11	8.4
Secondary Tr1 rest	16.6	Lung Microvascular EC none	22.5
Primary Th1 act	94.0	Lung Microvascular EC TNFalpha + IL-1beta	39.2
Primary Th2 act	66.4	Microvascular Dermal EC none	20.2
Primary Tr1 act	94.6	Microvascular Dermal EC TNFalpha + IL-1beta	30.4
Primary Th1 rest	53.6	Bronchial epithelium TNFalpha + IL1beta	9.2
Primary Th2 rest	30.1	Small airway epithelium none	6.1
Primary Tr1 rest	18.8	Small airway epithelium TNFalpha + IL-1beta	53.6
CD45RA CD4 lymphocyte act	24.7	Coronary artery SMC rest	16.7
CD45RO CD4 lymphocyte act	47.6	Coronary artery SMC TNFalpha + IL-1beta	4.6
CD8 lymphocyte act	20.4	Astrocytes rest	4.8
Secondary CD8 lymphocyte rest	24.8	Astrocytes TNFalpha + IL-1beta	16.2
Secondary CD8 lymphocyte act	27.5	KU-812 (Basophil) rest	6.6
CD4 lymphocyte none	9.9	KU-812 (Basophil) PMA/ionomycin	17.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	27.5	CCD1106 (Keratinocytes) none	2.1
LAK cells rest	28.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	4.2
LAK cells IL-2	33.2	Liver cirrhosis	5.6
LAK cells IL-2+IL-12	34.9	Lupus kidney	6.3
LAK cells IL-2+IFN gamma	47.3	NCI-H292 none	16.2
LAK cells IL-2+ IL-18	94.0	NCI-H292 IL-4	13.4
LAK cells	48.3	NCI-H292 IL-9	12.9

PMA/ionomycin			
NK Cells IL-2 rest	28.1	NCI-H292 IL-13	6.2
Two Way MLR 3 day	53.2	NCI-H292 IFN gamma	13.6
Two Way MLR 5 day	20.9	HPAEC none	14.6
Two Way MLR 7 day	22.1	HPAEC TNF alpha + IL-1 beta	35.1
PBMC rest	8.4	Lung fibroblast none	3.6
PBMC PWM	100.0	Lung fibroblast TNF alpha + IL-1 beta	6.4
PBMC PHA-L	34.9	Lung fibroblast IL-4	5.8
Ramos (B cell) none	4.5	Lung fibroblast IL-9	4.0
Ramos (B cell) ionomycin	13.0	Lung fibroblast IL-13	3.0
B lymphocytes PWM	96.6	Lung fibroblast IFN gamma	8.2
B lymphocytes CD40L and IL-4	84.7	Dermal fibroblast CCD1070 rest	9.5
EOL-1 dbcAMP	19.5	Dermal fibroblast CCD1070 TNF alpha	99.3
EOL-1 dbcAMP PMA/ionomycin	74.7	Dermal fibroblast CCD1070 IL-1 beta	15.2
Dendritic cells none	13.4	Dermal fibroblast IFN gamma	7.5
Dendritic cells LPS	52.9	Dermal fibroblast IL-4	6.7
Dendritic cells anti-CD40	12.0	IBD Colitis 2	1.3
Monocytes rest	22.2	IBD Crohn's	3.7
Monocytes LPS	67.8	Colon	15.7
Macrophages rest	18.9	Lung	11.3
Macrophages LPS	99.3	Thymus	28.3
HUVEC none	15.3	Kidney	31.9
HUVEC starved	34.2		

AI_comprehensive_panel_v1.0 Summary: Ag4520 The NOV93 gene is widely expressed among the samples on this panel, with highest expression in normal colon adjacent to diseased colon (CT=29). This widespread pattern of expression is consistent with expression in Panels 4D and 4.1D. Please see Panel 4.1D for discussion of utility of this gene in inflammation.

CNS_neurodegeneration_v1.0 Summary: Ag2904/Ag4520 The NOV93 gene, an IMP dehydrogenase homolog, shows a small but significant ($p=0.02$) upregulation in the postmortem Alzheimer's brain when compared to nondemented controls. IMP dehydrogenase is involved in purine metabolism, and has been implicated as a drug target for suppressing the

immune response, inflammation, and cerebral edema. The observed increase in the expression of this gene is in concordance with the evidence for the role of neuroinflammation in Alzheimer's disease. Therefore, the inhibition of this molecule may be of therapeutic benefit in Alzheimer's disease, head or spinal cord trauma, stroke, cerebral edema, or viral infections of the CNS.

References:

Hall IH, Wyrick SD. Cytotoxicity of [(5,6-dichloro-9a-n-propyl-2,3,9,9a-tetrahydro-3-oxo-1H fluoren-7-yl)oxy]acetic acid, an agent known to reduce brain edema. *Biomed Pharmacother* 1996;50(1):19-23

10 A known agent, [(5,6-dichloro-9a-n-propyl-2,3,9,9a-tetrahydro-3-oxo-1H fluoren-7-yl)oxy]acetic acid, which blocks brain edema, was also shown to be a potent cytotoxic agent in leukemia cells. The major site of action of the agents appears to be in the de novo purine synthetic pathway in L1210 leukemic cells. Both PRPP amido transferase and IMP dehydrogenase activities were suppressed by the agent. The inhibition of both regulatory
15 enzymes of the pathway along with the reduction of dihydrofolate reductase activity would account for the observed suppression of DNA and RNA syntheses and subsequent cancer cell death.

Senda M, Natsumeda Y. Tissue-differential expression of two distinct genes for human IMP dehydrogenase (E.C.1.1.1.205). *Life Sci* 1994;54(24):1917-26

20 Human IMP dehydrogenase (E.C. 1.1.1.205) is recently regarded as a potent targeting enzyme for immunosuppressive drugs. Tissue differential expressions of human type I and type II IMP dehydrogenase were investigated in sixteen human adult organs (heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocytes) and five human fetal organs (heart, brain,
25 lung, liver, kidney) using Northern blot analysis. In all tissues examined in this study, the sizes of mRNAs of each isoform were identical, respectively. The 2.3 kb type II mRNA was shown predominantly, and the 3.5 kb type I mRNA level was lower than type II in most human tissues examined. In contrast, type I IMPDH gene expressed higher than type II in peripheral blood leukocytes, uniquely. We also demonstrated that both type I and type II IMPDH genes
30 are widely distributed among various species by Southern blot analysis. Interestingly, type I IMPDH gene may have multiple gene families in primates. [dstone, 17-Jan-02]

General_screening_panel_v1.4 Summary: Ag4250 Two experiments with the same probe and primer set produce results that are in excellent agreement, with highest expression of the NOV93 gene in a gastric cancer cell line and fetal lung tissue. In addition, there appears

with normal lung tissue when compared to adjacent lung cancer tissue. Thus, the expression of this gene could be used to distinguish this sample of normal lung tissue from other samples in the panel. In addition, the expression of this gene could be used to distinguish normal lung tissue adjacent to cancer tissue. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, antibodies or protein therapeutics might be beneficial in the treatment of lung cancer.

Panel 4D/4.1D Summary: Ag2904/Ag4520 The NOV93 gene, a novel IMP dehydrogenase-like protein, is differentially expressed, as displayed in Panels 4.1D and 4D, in activated T cells, activated B cells, activated monocytes, activated macrophages, and activated dendritic cells. Small molecule antagonists of the previously characterized IMP dehydrogenase have been found to be useful in the treatment of several immunopathological states (See Allison and Eugui, 2001). Therefore, small molecule antagonists of the NOV93 gene product may reduce or eliminate the symptoms of autoimmune and inflammatory diseases, including Crohn's disease, ulcerative colitis, multiple sclerosis, chronic obstructive pulmonary disease, asthma, emphysema, rheumatoid arthritis, lupus erythematosus, or psoriasis.

References:

Allison AC, Eugui EM. Mycophenolate mofetil and its mechanisms of action. Immunopharmacology 2000 May;47(2-3):85-118

Mycophenolate mofetil (MMF, CellCept(R)) is a prodrug of mycophenolic acid (MPA), an inhibitor of inosine monophosphate dehydrogenase (IMPDH). This is the rate-limiting enzyme in de novo synthesis of guanosine nucleotides. T- and B-lymphocytes are more dependent on this pathway than other cell types are. Moreover, MPA is a fivefold more potent inhibitor of the type II isoform of IMPDH, which is expressed in activated lymphocytes, than of the type I isoform of IMPDH, which is expressed in most cell types. MPA has therefore a more potent cytostatic effect on lymphocytes than on other cell types. This is the principal mechanism by which MPA exerts immunosuppressive effects. Three other mechanisms may also contribute to the efficacy of MPA in preventing allograft rejection and other applications. First, MPA can induce apoptosis of activated T-lymphocytes, which may eliminate clones of cells responding to antigenic stimulation. Second, by depleting guanosine nucleotides, MPA suppresses glycosylation and the expression of some adhesion molecules, thereby decreasing the recruitment of lymphocytes and monocytes into sites of inflammation and graft rejection. Third, by depleting guanosine nucleotides MPA also depletes tetrahydrobiopterin, a co-factor for the inducible form of nitric oxide synthase (iNOS). MPA therefore suppresses the production by iNOS of NO, and consequent tissue damage mediated

by peroxyxynitrite. CellCept(R) suppresses T-lymphocytic responses to allogeneic cells and other antigens. The drug also suppresses primary, but not secondary, antibody responses. The efficacy of regimes including CellCept(R) in preventing allograft rejection, and in the treatment of rejection, is now firmly established. CellCept(R) is also efficacious in several experimental animal models of chronic rejection, and it is hoped that the drug will have the same effect in humans.

NOV94

Expression of gene NOV94 was assessed using the primer-probe set Ag2905, described in Table CIA. Results of the RTQ-PCR runs are shown in Tables CIB, CIC and CID.

Table CIA. Probe Name Ag2905

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gcagaataccacgatgacttct-3'	22	1468	1366
Probe	TET-5'-agtcagcttacgtcgctgcctctgag-3'-TAMRA	26	1496	1367
Reverse	5'-gttctctggtgctgtaatgca-3'	20	1523	1368

Table CIB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag2905, Run 161374149	Tissue Name	Rel. Exp.(%) Ag2905, Run 161374149
Liver adenocarcinoma	20.4	Kidney (fetal)	1.4
Pancreas	0.1	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	4.9	Renal ca. A498	7.4
Adrenal gland	0.2	Renal ca. RXF 393	9.0
Thyroid	0.9	Renal ca. ACHN	2.4
Salivary gland	0.6	Renal ca. UO-31	5.5
Pituitary gland	0.4	Renal ca. TK-10	13.7
Brain (fetal)	0.4	Liver	0.3
Brain (whole)	0.2	Liver (fetal)	0.8
Brain (amygdala)	1.2	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.2	Lung	0.6
Brain (hippocampus)	1.0	Lung (fetal)	0.9
Brain (substantia nigra)	0.3	Lung ca. (small cell) LX-1	15.9
Brain (thalamus)	1.0	Lung ca. (small cell) NCI-H69	1.2

Cerebral Cortex	2.3	Lung ca. (s.cell var.) SHP-77	27.7
Spinal cord	1.5	Lung ca. (large cell)NCI-H460	2.0
glio/astro U87-MG	11.7	Lung ca. (non-sm. cell) A549	8.0
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	2.8
astrocytoma SW1783	0.2	Lung ca. (non-s.cell) HOP-62	1.9
neuro*; met SK-N-AS	5.8	Lung ca. (non-s.cl) NCI-H522	5.4
astrocytoma SF-539	5.3	Lung ca. (squam.) SW 900	2.6
astrocytoma SNB-75	1.7	Lung ca. (squam.) NCI-H596	1.6
glioma SNB-19	0.1	Mammary gland	2.2
glioma U251	0.0	Breast ca.* (pl.ef) MCF-7	40.6
glioma SF-295	0.3	Breast ca.* (pl.ef) MDA-MB-231	6.3
Heart (fetal)	0.2	Breast ca.* (pl.ef) T47D	0.0
Heart	10.2	Breast ca. BT-549	1.2
Skeletal muscle (fetal)	3.4	Breast ca. MDA-N	10.1
Skeletal muscle	0.7	Ovary	1.9
Bone marrow	1.0	Ovarian ca. OVCAR- 3	1.6
Thymus	20.2	Ovarian ca. OVCAR- 4	0.3
Spleen	0.8	Ovarian ca. OVCAR- 5	7.8
Lymph node	0.2	Ovarian ca. OVCAR- 8	10.2
Colorectal	1.0	Ovarian ca. IGROV- 1	1.5
Stomach	1.9	Ovarian ca.* (ascites) SK-OV-3	9.4
Small intestine	2.3	Uterus	0.1
Colon ca. SW480	6.0	Placenta	1.6
Colon ca.* SW620(SW480 met)	9.2	Prostate	3.5
Colon ca. HT29	30.8	Prostate ca.* (bone met)PC-3	9.3
Colon ca. HCT-116	11.2	Testis	100.0

Colon ca. CaCo-2	11.3	Melanoma Hs688(A).T	1.2
Colon ca. tissue(ODO3866)	2.4	Melanoma* (met) Hs688(B).T	2.0
Colon ca. HCC-2998	7.2	Melanoma UACC-62	1.5
Gastric ca.* (liver met) NCI-N87	40.6	Melanoma M14	1.0
Bladder	2.0	Melanoma LOX IMVI	0.6
Trachea	4.0	Melanoma* (met) SK-MEL-5	1.0
Kidney	1.0	Adipose	0.4

Table CIC. Panel 2D

Tissue Name	Rel. Exp.(%) Ag2905, Run 161374481	Tissue Name	Rel. Exp.(%) Ag2905, Run 161374481
Normal Colon	15.3	Kidney Margin 8120608	1.1
CC Well to Mod Diff (ODO3866)	5.4	Kidney Cancer 8120613	0.5
CC Margin (ODO3866)	1.1	Kidney Margin 8120614	0.0
CC Gr.2 rectosigmoid (ODO3868)	3.1	Kidney Cancer 9010320	42.3
CC Margin (ODO3868)	1.1	Kidney Margin 9010321	2.0
CC Mod Diff (ODO3920)	26.6	Normal Uterus	1.8
CC Margin (ODO3920)	3.6	Uterus Cancer 064011	8.2
CC Gr.2 ascend colon (ODO3921)	14.2	Normal Thyroid	9.2
CC Margin (ODO3921)	6.2	Thyroid Cancer 064010	16.0
CC from Partial Hepatectomy (ODO4309) Mets	44.8	Thyroid Cancer A302152	10.2
Liver Margin (ODO4309)	8.4	Thyroid Margin A302153	15.1
Colon mets to lung (OD04451-01)	11.0	Normal Breast	9.7
Lung Margin (OD04451- 02)	0.9	Breast Cancer (OD04566)	1.2
Normal Prostate 6546-1	10.1	Breast Cancer (OD04590-01)	3.5
Prostate Cancer	52.9	Breast Cancer Mets	2.0

(OD04410)		(OD04590-03)	
Prostate Margin (OD04410)	23.0	Breast Cancer Metastasis (OD04655-05)	54.7
Prostate Cancer (OD04720-01)	18.7	Breast Cancer 064006	8.1
Prostate Margin (OD04720-02)	33.2	Breast Cancer 1024	26.1
Normal Lung 061010	12.9	Breast Cancer 9100266	30.6
Lung Met to Muscle (ODO4286)	29.3	Breast Margin 9100265	17.4
Muscle Margin (ODO4286)	2.7	Breast Cancer A209073	62.0
Lung Malignant Cancer (OD03126)	3.4	Breast Margin A2090734	20.7
Lung Margin (OD03126)	7.5	Normal Liver	11.7
Lung Cancer (OD04404)	35.6	Liver Cancer 064003	6.7
Lung Margin (OD04404)	4.2	Liver Cancer 1025	3.4
Lung Cancer (OD04565)	0.7	Liver Cancer 1026	8.2
Lung Margin (OD04565)	2.2	Liver Cancer 6004-T	3.3
Lung Cancer (OD04237- 01)	51.8	Liver Tissue 6004-N	15.3
Lung Margin (OD04237- 02)	0.4	Liver Cancer 6005-T	10.4
Ocular Mel Met to Liver (ODO4310)	41.8	Liver Tissue 6005-N	0.2
Liver Margin (ODO4310)	4.4	Normal Bladder	16.7
Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer 1023	6.3
Lung Margin (OD04321)	4.7	Bladder Cancer A302173	40.6
Normal Kidney	7.1	Bladder Cancer (OD04718-01)	1.0
Kidney Ca, Nuclear grade 2 (OD04338)	16.5	Bladder Normal Adjacent (OD04718- 03)	5.7
Kidney Margin (OD04338)	7.4	Normal Ovary	1.3
Kidney Ca Nuclear grade 1/2 (OD04339)	12.9	Ovarian Cancer 064008	100.0
Kidney Margin (OD04339)	4.6	Ovarian Cancer (OD04768-07)	18.3
Kidney Ca, Clear cell type (OD04340)	5.8	Ovary Margin (OD04768-08)	0.7

Kidney Margin (OD04340)	3.5	Normal Stomach	5.8
Kidney Ca, Nuclear grade 3 (OD04348)	4.0	Gastric Cancer 9060358	0.9
Kidney Margin (OD04348)	4.4	Stomach Margin 9060359	4.8
Kidney Cancer (OD04622-01)	0.7	Gastric Cancer 9060395	13.2
Kidney Margin (OD04622-03)	1.3	Stomach Margin 9060394	8.3
Kidney Cancer (OD04450-01)	11.9	Gastric Cancer 9060397	14.7
Kidney Margin (OD04450-03)	8.8	Stomach Margin 9060396	2.1
Kidney Cancer 8120607	0.0	Gastric Cancer 064005	17.1

Table CID. Panel 4D

Tissue Name	Rel. Exp.(%) Ag2905, Run 159772697	Tissue Name	Rel. Exp.(%) Ag2905, Run 159772697
Secondary Th1 act	12.8	HUVEC IL-1beta	8.2
Secondary Th2 act	6.5	HUVEC IFN gamma	9.1
Secondary Tr1 act	8.7	HUVEC TNF alpha + IFN gamma	11.6
Secondary Th1 rest	4.7	HUVEC TNF alpha + IL4	12.2
Secondary Th2 rest	6.0	HUVEC IL-11	9.0
Secondary Tr1 rest	6.7	Lung Microvascular EC none	1.4
Primary Th1 act	12.5	Lung Microvascular EC TNFalpha + IL-1beta	3.1
Primary Th2 act	10.6	Microvascular Dermal EC none	24.3
Primary Tr1 act	16.6	Microvascular Dermal EC TNFalpha + IL-1beta	13.7
Primary Th1 rest	24.5	Bronchial epithelium TNFalpha + IL1beta	1.0
Primary Th2 rest	10.2	Small airway epithelium none	6.7
Primary Tr1 rest	17.4	Small airway epithelium TNFalpha + IL-1beta	29.1
CD45RA CD4 lymphocyte act	3.8	Coronary artery SMC rest	9.9
CD45RO CD4 lymphocyte act	8.9	Coronary artery SMC TNFalpha + IL-1beta	6.0

CD8 lymphocyte act	5.1	Astrocytes rest	9.1
Secondary CD8 lymphocyte rest	8.2	Astrocytes TNFalpha + IL-1beta	5.4
Secondary CD8 lymphocyte act	7.6	KU-812 (Basophil) rest	10.7
CD4 lymphocyte none	1.0	KU-812 (Basophil) PMA/ionomycin	21.6
2ry Th1/Th2/Tr1_anti-CD95 CH11	8.9	CCD1106 (Keratinocytes) none	30.4
LAK cells rest	4.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	1.1
LAK cells IL-2	10.0	Liver cirrhosis	1.1
LAK cells IL-2+IL-12	11.2	Lupus kidney	1.0
LAK cells IL-2+IFN gamma	14.7	NCI-H292 none	35.4
LAK cells IL-2+ IL-18	11.7	NCI-H292 IL-4	29.7
LAK cells PMA/ionomycin	2.8	NCI-H292 IL-9	49.0
NK Cells IL-2 rest	5.0	NCI-H292 IL-13	17.9
Two Way MLR 3 day	5.3	NCI-H292 IFN gamma	32.8
Two Way MLR 5 day	5.8	HPAEC none	4.8
Two Way MLR 7 day	3.4	HPAEC TNF alpha + IL-1 beta	5.0
PBMC rest	0.8	Lung fibroblast none	7.3
PBMC PWM	14.4	Lung fibroblast TNF alpha + IL-1 beta	3.9
PBMC PHA-L	9.0	Lung fibroblast IL-4	8.4
Ramos (B cell) none	17.4	Lung fibroblast IL-9	10.9
Ramos (B cell) ionomycin	100.0	Lung fibroblast IL-13	8.8
B lymphocytes PWM	31.2	Lung fibroblast IFN gamma	17.1
B lymphocytes CD40L and IL-4	15.5	Dermal fibroblast CCD1070 rest	31.9
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	39.2
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast CCD1070 IL-1 beta	12.2
Dendritic cells none	4.9	Dermal fibroblast IFN gamma	6.4
Dendritic cells LPS	3.0	Dermal fibroblast IL-4	18.2
Dendritic cells anti-CD40	4.9	IBD Colitis 2	0.7
Monocytes rest	2.0	IBD Crohn's	0.0
Monocytes LPS	2.0	Colon	5.4

Macrophages rest	3.6	Lung	5.2
Macrophages LPS	1.8	Thymus	3.0
HUVEC none	19.2	Kidney	20.0
HUVEC starved	24.0		

Panel 1.3D Summary: Ag2905 The expression of the NOV94 gene appears to be highest in a sample derived from normal testis tissue (CT=28.9). In addition, there is substantial expression associated with samples derived from colon cancer cell lines, lung cancer cell lines and breast cancer cell lines. Thus, the expression of this gene could be used to distinguish normal testis tissue from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies could be beneficial for the treatment of colon, lung or breast cancer.

In addition, this gene appears to be differentially expressed in fetal (CT value = 37) vs adult heart (CT value = 32), and may be useful for the differentiation between the two sources of heart tissue.

Panel 2D Summary: Ag2905 The expression of the NOV94 gene appears to be highest in a sample derived from an ovarian cancer (CT=30.5). In addition, there appears to be substantial expression associated with breast cancers, lung cancers, gastric cancers, prostate cancers and colon cancers. Thus, the expression of this gene could be used to distinguish this ovarian cancer sample from others in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be beneficial in the treatment of ovarian, breast, lung, gastric, prostate or colon cancer.

Panel 4D Summary: Ag2905 Low but significant expression of the NOV94 transcript is found predominantly in activated B cell lymphoma cell line (Ramos) and activated B cells (CTs=32-34). It is also found in HUVEC, keratinocytes, lung fibroblasts and the mucoid epidermoid cell line H292. Therefore, targeting of this gene product with a small molecule drug therapeutic may modulate the functions of B cells and lead to the improvement of symptoms in autoimmune diseases such as lupus erythematosus, rheumatoid arthritis, hyperglobulinemia and other B cell disorders.

NOV95

Expression of gene NOV95 was assessed using the primer-probe set Ag3060, described in Table CJA. Results of the RTQ-PCR runs are shown in Tables CJB and CJC.

Table CJA. Probe Name Ag3060

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-caaagattgcagcaatcgatag-3'	22	189	1369
Probe	TET-5'-agtatacacgaggctttggccatcca-3'- TAMRA	26	219	1370
Reverse	5'-aggacagagctttcacaagtga-3'	22	245	1371

Table CJB. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3060, Run 168016485	Tissue Name	Rel. Exp.(%) Ag3060, Run 168016485
Liver adenocarcinoma	0.2	Kidney (fetal)	0.3
Pancreas	0.2	Renal ca. 786-0	0.1
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.1
Adrenal gland	0.1	Renal ca. RXF 393	0.1
Thyroid	0.0	Renal ca. ACHN	0.1
Salivary gland	0.0	Renal ca. UO-31	0.1
Pituitary gland	0.1	Renal ca. TK-10	0.1
Brain (fetal)	0.1	Liver	0.0
Brain (whole)	0.1	Liver (fetal)	0.0
Brain (amygdala)	0.1	Liver ca. (hepatoblast) HepG2	0.1
Brain (cerebellum)	0.1	Lung	0.0
Brain (hippocampus)	0.2	Lung (fetal)	0.1
Brain (substantia nigra)	0.1	Lung ca. (small cell) LX-1	0.1
Brain (thalamus)	0.1	Lung ca. (small cell) NCI-H69	0.1
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.2
Spinal cord	0.1	Lung ca. (large cell)NCI-H460	0.0
glio/astro U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.3
glio/astro U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.1
astrocytoma SW1783	0.2	Lung ca. (non-s.cell) HOP-62	0.1
neuro*; met SK-N-AS	0.1	Lung ca. (non-s.cl) NCI-H522	0.2
astrocytoma SF-539	0.1	Lung ca. (squam.) SW 900	0.1
astrocytoma SNB-75	0.2	Lung ca. (squam.)	0.1

		NCI-H596	
glioma SNB-19	0.1	Mammary gland	0.1
glioma U251	0.3	Breast ca.* (pl.ef) MCF-7	0.1
glioma SF-295	0.2	Breast ca.* (pl.ef) MDA-MB-231	0.1
Heart (fetal)	3.3	Breast ca.* (pl.ef) T47D	100.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal muscle (fetal)	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.2
Thymus	0.1	Ovarian ca. OVCAR-4	0.2
Spleen	0.1	Ovarian ca. OVCAR-5	0.6
Lymph node	0.1	Ovarian ca. OVCAR-8	0.1
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.2
Small intestine	0.0	Uterus	0.1
Colon ca. SW480	0.1	Placenta	0.0
Colon ca.* SW620(SW480 met)	0.2	Prostate	0.0
Colon ca. HT29	0.1	Prostate ca.* (bone met)PC-3	0.1
Colon ca. HCT-116	0.1	Testis	3.1
Colon ca. CaCo-2	0.1	Melanoma Hs688(A).T	0.0
Colon ca. tissue(ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.1
Colon ca. HCC-2998	0.2	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.1	Melanoma M14	0.1
Bladder	0.1	Melanoma LOX IMVI	0.0
Trachea	0.1	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.1	Adipose	0.1

Table CJC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3060, Run 164317425	Tissue Name	Rel. Exp.(%) Ag3060, Run 164317425
Secondary Th1 act	22.7	HUVEC IL-1beta	7.2
Secondary Th2 act	25.7	HUVEC IFN gamma	15.1
Secondary Tr1 act	40.3	HUVEC TNF alpha + IFN gamma	12.3
Secondary Th1 rest	6.1	HUVEC TNF alpha + IL4	10.4
Secondary Th2 rest	9.3	HUVEC IL-11	4.8
Secondary Tr1 rest	11.7	Lung Microvascular EC none	6.0
Primary Th1 act	29.5	Lung Microvascular EC TNFalpha + IL-1beta	9.7
Primary Th2 act	24.1	Microvascular Dermal EC none	11.8
Primary Tr1 act	33.4	Microsvascular Dermal EC TNFalpha + IL-1beta	9.6
Primary Th1 rest	52.9	Bronchial epithelium TNFalpha + IL1beta	20.0
Primary Th2 rest	26.2	Small airway epithelium none	5.0
Primary Tr1 rest	17.4	Small airway epithelium TNFalpha + IL-1beta	38.4
CD45RA CD4 lymphocyte act	13.8	Coronary artery SMC rest	12.1
CD45RO CD4 lymphocyte act	23.0	Coronary artery SMC TNFalpha + IL-1beta	8.7
CD8 lymphocyte act	26.4	Astrocytes rest	9.0
Secondary CD8 lymphocyte rest	25.9	Astrocytes TNFalpha + IL-1beta	5.6
Secondary CD8 lymphocyte act	15.1	KU-812 (Basophil) rest	22.7
CD4 lymphocyte none	7.1	KU-812 (Basophil) PMA/ionomycin	67.8
2ry Th1/Th2/Tr1_anti- CD95 CH11	13.4	CCD1106 (Keratinocytes) none	10.3
LAK cells rest	15.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	5.8
LAK cells IL-2	18.7	Liver cirrhosis	2.6
LAK cells IL-2+IL-12	14.5	Lupus kidney	1.0
LAK cells IL-2+IFN gamma	24.1	NCI-H292 none	22.1
LAK cells IL-2+ IL-18	19.5	NCI-H292 IL-4	21.8
LAK cells	11.7	NCI-H292 IL-9	28.7

PMA/ionomycin			
NK Cells IL-2 rest	14.2	NCI-H292 IL-13	15.5
Two Way MLR 3 day	14.0	NCI-H292 IFN gamma	27.4
Two Way MLR 5 day	14.5	HPAEC none	7.1
Two Way MLR 7 day	13.7	HPAEC TNF alpha + IL-1 beta	11.5
PBMC rest	4.2	Lung fibroblast none	8.9
PBMC PWM	38.2	Lung fibroblast TNF alpha + IL-1 beta	10.8
PBMC PHA-L	21.3	Lung fibroblast IL-4	14.6
Ramos (B cell) none	14.4	Lung fibroblast IL-9	13.9
Ramos (B cell) ionomycin	67.8	Lung fibroblast IL-13	11.0
B lymphocytes PWM	100.0	Lung fibroblast IFN gamma	20.9
B lymphocytes CD40L and IL-4	27.2	Dermal fibroblast CCD1070 rest	16.3
EOL-1 dbcAMP	13.4	Dermal fibroblast CCD1070 TNF alpha	32.3
EOL-1 dbcAMP PMA/ionomycin	14.8	Dermal fibroblast CCD1070 IL-1 beta	7.3
Dendritic cells none	11.4	Dermal fibroblast IFN gamma	10.2
Dendritic cells LPS	15.9	Dermal fibroblast IL-4	18.2
Dendritic cells anti-CD40	10.8	IBD Colitis 2	0.5
Monocytes rest	10.6	IBD Crohn's	1.2
Monocytes LPS	7.0	Colon	6.8
Macrophages rest	12.2	Lung	9.5
Macrophages LPS	10.9	Thymus	14.3
HUVEC none	7.0	Kidney	28.9
HUVEC starved	17.7		

Panel 1.3D Summary: Ag3060 Results from one experiment with this gene are not included. The amp plot indicates that there were experimental difficulties with this run (data not shown).

5 **Panel 4D Summary:** Ag3060 High expression of the NOV95 transcript (CTs= 26.3-26.9) is found in activated B cells and B cell lymphoma (Ramos). B cells generate antibody response and lead to activation of T cell mediated response as antigen presenting cells and are central to the function of the immune response. Therefore, targeting of this gene product with a small molecule drug therapeutic may modulate the functions of B cells and lead to the

improvement of symptoms of autoimmune diseases such as lupus erythematosus, rheumatoid arthritis, hyperglobulinemia and other B cell disorders.

In addition, moderate expression of this gene is also found in a wide range of cell types of significance in the immune response in health and diseases. This suggests the broader involvement of the protein encoded by this gene in many inflammatory and autoimmune diseases.

NOV96a, NOV96b, and NOV96c

Expression of gene NOV96a and full length clones NOV96b and NOV96c was assessed using the primer-probe set Ag4532, described in Table CKA. Results of the RTQ-PCR runs are shown in Table CKB.

Table CKA. Probe Name Ag4532

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-actcacctctctcctccatcat-3'	22	626	1372
Probe	TET-5'-cgttacactggtgcctcaccctgat-3'-TAMRA	26	660	1373
Reverse	5'-aggggaatgaagtagccagtgtt-3'	22	687	1374

Table CKB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag4532, Run 222735297	Tissue Name	Rel. Exp.(%) Ag4532, Run 222735297
Adipose	2.0	Renal ca. TK-10	31.9
Melanoma* Hs688(A).T	3.6	Bladder	3.7
Melanoma* Hs688(B).T	5.0	Gastric ca. (liver met.) NCI-N87	26.6
Melanoma* M14	18.2	Gastric ca. KATO III	33.7
Melanoma* LOXIMVI	5.4	Colon ca. SW-948	15.0
Melanoma* SK-MEL-5	2.7	Colon ca. SW480	31.2
Squamous cell carcinoma SCC-4	6.3	Colon ca.* (SW480 met) SW620	18.4
Testis Pool	1.0	Colon ca. HT29	4.2
Prostate ca.* (bone met) PC-3	7.3	Colon ca. HCT-116	20.6
Prostate Pool	1.5	Colon ca. CaCo-2	17.1
Placenta	9.0	Colon cancer tissue	20.4
Uterus Pool	0.8	Colon ca. SW1116	6.0

Ovarian ca. OVCAR-3	10.7	Colon ca. Colo-205	12.3
Ovarian ca. SK-OV-3	2.5	Colon ca. SW-48	13.1
Ovarian ca. OVCAR-4	9.0	Colon Pool	3.4
Ovarian ca. OVCAR-5	29.5	Small Intestine Pool	2.7
Ovarian ca. IGROV-1	4.0	Stomach Pool	1.1
Ovarian ca. OVCAR-8	14.6	Bone Marrow Pool	1.3
Ovary	1.7	Fetal Heart	1.7
Breast ca. MCF-7	17.1	Heart Pool	2.1
Breast ca. MDA-MB-231	20.0	Lymph Node Pool	2.7
Breast ca. BT 549	32.3	Fetal Skeletal Muscle	0.7
Breast ca. T47D	68.8	Skeletal Muscle Pool	1.2
Breast ca. MDA-N	13.5	Spleen Pool	5.2
Breast Pool	2.3	Thymus Pool	4.0
Trachea	7.3	CNS cancer (glio/astro) U87-MG	47.3
Lung	0.2	CNS cancer (glio/astro) U-118-MG	4.0
Fetal Lung	8.4	CNS cancer (neuro;met) SK-N-AS	1.4
Lung ca. NCI-N417	1.0	CNS cancer (astro) SF-539	13.6
Lung ca. LX-1	21.2	CNS cancer (astro) SNB-75	16.5
Lung ca. NCI-H146	0.2	CNS cancer (glio) SNB-19	4.4
Lung ca. SHP-77	13.8	CNS cancer (glio) SF-295	9.3
Lung ca. A549	44.4	Brain (Amygdala) Pool	0.5
Lung ca. NCI-H526	3.1	Brain (cerebellum)	4.2
Lung ca. NCI-H23	19.5	Brain (fetal)	3.8
Lung ca. NCI-H460	7.5	Brain (Hippocampus) Pool	1.1
Lung ca. HOP-62	17.4	Cerebral Cortex Pool	0.5
Lung ca. NCI-H522	26.6	Brain (Substantia nigra) Pool	1.1
Liver	23.3	Brain (Thalamus) Pool	1.3
Fetal Liver	34.2	Brain (whole)	1.3
Liver ca. HepG2	32.8	Spinal Cord Pool	1.3

Kidney Pool	4.6	Adrenal Gland	2.8
Fetal Kidney	3.8	Pituitary gland Pool	0.6
Renal ca. 786-0	9.3	Salivary Gland	6.3
Renal ca. A498	10.2	Thyroid (female)	6.0
Renal ca. ACHN	100.0	Pancreatic ca. CAPAN2	32.1
Renal ca. UO-31	22.2	Pancreas Pool	2.4

General_screening_panel_v1.4 Summary: Ag4532 The expression of the NOV96a gene appears to be highest in a sample derived from a renal cancer cell line

(ACHN)(CT=26.4). In addition, there is substantial expression associated with other renal

5 cancer cell lines as well as gastric cancer cell lines, colon cancer cell lines, lung cancer cell lines, and breast cancer cell lines. Thus, the expression of this gene could be used to

distinguish ACHN cells from other samples in this panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics or antibodies might be of benefit in the treatment of kidney, gastric, colon, lung or breast cancer.

10 Among metabolic tissues, this gene has low-to-moderate levels of expression in adipose, liver, heart, skeletal muscle, adrenal, pituitary, thyroid, and pancreas. Thus, this gene product may be a small molecule target for the treatment of endocrine and metabolic diseases, including obesity and Types 1 and 2 diabetes. The direction of therapeutic modulation for this gene product would, of necessity, be tissue- or organ-specific. The consequences of altered

15 lactate/monocarboxylate/ketone body transport would differ dramatically between tissues.

In addition, this gene, a monocarboxylate transporter homolog, is expressed at low to moderate levels in all CNS regions examined. The monocarboxylate transporters have been implicated in post-ischemic neuronal loss in stroke, such that blockade of these transporters increase stroke-related damage. Thus, this gene is an excellent drug target, such that increasing

20 its activity may decrease postischemic damage in stroke/cerebral infarct.

References:

Schurr A, Payne RS, Miller JJ, Tseng MT, Rigor BM. Blockade of lactate transport exacerbates delayed neuronal damage in a rat model of cerebral ischemia. Brain Res 2001 Mar 23;895(1-2):268-72

25 Studies over the past decade have demonstrated that lactate is produced aerobically during brain activation and it has been suggested to be an obligatory aerobic energy substrate postischemia. It has been also hypothesized, based on in vitro studies, that lactate, produced by glia in large amounts during activation and/or ischemia/hypoxia, is transported via specific glial and neuronal monocarboxylate transporters into neurons for aerobic utilization. To test

the role of lactate as an aerobic energy substrate postischemia in vivo, we employed the cardiac-arrest-induced transient global cerebral ischemia (TGI) rat model and the monocarboxylate transporter inhibitor alpha-cyano-4-hydroxycinnamate (4-CIN). Once 4-CIN was established to cross the blood-brain barrier, rats were treated with the inhibitor 60 min prior to a 5-min TGI. These rats exhibited a significantly greater degree of delayed neuronal damage in the hippocampus than control, untreated rats, as measured 7 days post-TGI. We concluded that intra-ischemically-accumulated lactate is utilized aerobically as the main energy substrate immediately postischemia. Blockade of lactate transport into neurons prevents its utilization and, consequently, exacerbates delayed ischemic neuronal damage.

10 NOV97c and NOV97d

Expression of gene NOV97c and variant NOV97d was assessed using the primer-probe set Ag3697, described in Table CLA. Results of the RTQ-PCR runs are shown in Table CLB.

Table CLA. Probe Name Ag3697

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-catctggattgacactggaatt-3'	22	585	1375
Probe	TET-5'-actcccgaggagtggatcacccat-3'-TAMRA	23	608	1376
Reverse	5'-aatcttattggcagtcacagatg-3'	22	639	1377

15

Table CLB. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3697, Run 218253566	Tissue Name	Rel. Exp.(%) Ag3697, Run 218253566
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.3	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	1.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	8.2
Squamous cell carcinoma SCC-4	0.9	Colon ca.* (SW480 met) SW620	6.9
Testis Pool	100.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.3	Colon ca. HCT-116	0.8
Prostate Pool	0.0	Colon ca. CaCo-2	0.3

Placenta	7.8	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.6	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	1.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	4.1	Colon Pool	0.7
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.8	Stomach Pool	0.0
Ovarian ca. OVCAR-8	2.0	Bone Marrow Pool	0.0
Ovary	0.2	Fetal Heart	0.0
Breast ca. MCF-7	0.5	Heart Pool	0.7
Breast ca. MDA-MB-231	1.2	Lymph Node Pool	0.2
Breast ca. BT 549	0.4	Fetal Skeletal Muscle	4.5
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.7
Breast ca. MDA-N	0.0	Spleen Pool	1.0
Breast Pool	0.3	Thymus Pool	1.8
Trachea	1.4	CNS cancer (glio/astro) U87-MG	1.1
Lung	0.9	CNS cancer (glio/astro) U-118-MG	4.7
Fetal Lung	1.2	CNS cancer (neuro;met) SK-N-AS	0.2
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	2.5	CNS cancer (astro) SNB-75	1.7
Lung ca. NCI-H146	1.4	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	28.9	CNS cancer (glio) SF-295	0.9
Lung ca. A549	0.5	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.2
Lung ca. NCI-H23	0.8	Brain (fetal)	0.0
Lung ca. NCI-H460	0.7	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	1.4	Cerebral Cortex Pool	0.5
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.3
Liver	0.0	Brain (Thalamus) Pool	1.6

Fetal Liver	0.5	Brain (whole)	0.7
Liver ca. HepG2	0.3	Spinal Cord Pool	0.3
Kidney Pool	0.0	Adrenal Gland	0.7
Fetal Kidney	0.0	Pituitary gland Pool	1.1
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.3	Pancreatic ca. CAPAN2	0.4
Renal ca. UO-31	4.1	Pancreas Pool	2.5

CNS_neurodegeneration_v1.0 Summary: Ag3697 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

General_screening_panel_v1.4 Summary: Ag3697 Expression of this gene is highest in and almost exclusive to testis (CT = 30.7). Therefore, expression of this gene could be used to distinguish testis from the other samples on this panel. Moreover, therapeutic modulation of the activity of this gene or its protein product using protein therapeutics, antibodies or small molecule drugs could be of benefit in the treatment of infertility.

Panel 4.1D Summary: Ag3697 Expression of this gene is low/undetectable (CTs > 35) across all of the samples on this panel (data not shown).

NOV98: AGRIN

Expression of gene NOV98 was assessed using the primer-probe set Ag3974, described in Table CMA. Results of the RTQ-PCR runs are shown in Tables CMB, CMC and CMD.

Table CMA. Probe Name Ag3974

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-gacaccaggatcttctttgtga-3'	22	274	1378
Probe	TET-5'-catacctgtggccagccacaag-3'-TAMRA	23	308	1379
Reverse	5'-gagttgagcatcagctcgtt-3'	20	331	1380

Table CMB. CNS_neurodegeneration_v1.0

Tissue Name	Rel. Exp.(%) Ag3974, Run 212348647	Tissue Name	Rel. Exp.(%) Ag3974, Run 212348647
AD 1 Hippo	28.7	Control (Path) 3 Temporal Ctx	15.4
AD 2 Hippo	36.3	Control (Path) 4 Temporal Ctx	46.0

AD 3 Hippo	19.2	AD 1 Occipital Ctx	28.9
AD 4 Hippo	21.8	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	80.7	AD 3 Occipital Ctx	19.5
AD 6 Hippo	42.6	AD 4 Occipital Ctx	21.0
Control 2 Hippo	42.3	AD 5 Occipital Ctx	47.6
Control 4 Hippo	34.9	AD 6 Occipital Ctx	14.7
Control (Path) 3 Hippo	12.4	Control 1 Occipital Ctx	20.4
AD 1 Temporal Ctx	32.1	Control 2 Occipital Ctx	55.1
AD 2 Temporal Ctx	31.2	Control 3 Occipital Ctx	22.5
AD 3 Temporal Ctx	20.2	Control 4 Occipital Ctx	22.2
AD 4 Temporal Ctx	24.0	Control (Path) 1 Occipital Ctx	71.2
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	17.4
AD 5 Sup Temporal Ctx	58.6	Control (Path) 3 Occipital Ctx	15.4
AD 6 Inf Temporal Ctx	40.6	Control (Path) 4 Occipital Ctx	38.7
AD 6 Sup Temporal Ctx	35.1	Control 1 Parietal Ctx	18.2
Control 1 Temporal Ctx	19.8	Control 2 Parietal Ctx	67.4
Control 2 Temporal Ctx	48.3	Control 3 Parietal Ctx	21.2
Control 3 Temporal Ctx	17.8	Control (Path) 1 Parietal Ctx	51.4
Control 3 Temporal Ctx	25.0	Control (Path) 2 Parietal Ctx	32.3
Control (Path) 1 Temporal Ctx	63.7	Control (Path) 3 Parietal Ctx	11.9
Control (Path) 2 Temporal Ctx	43.5	Control (Path) 4 Parietal Ctx	58.6

Table CMC. General_screening_panel_v1.4

Tissue Name	Rel. Exp.(%) Ag3974, Run 217508632	Tissue Name	Rel. Exp.(%) Ag3974, Run 217508632
Adipose	1.5	Renal ca. TK-10	16.4
Melanoma* Hs688(A).T	3.2	Bladder	9.0

Melanoma* Hs688(B).T	4.2	Gastric ca. (liver met.) NCI-N87	80.7
Melanoma* M14	6.4	Gastric ca. KATO III	17.7
Melanoma* LOXIMVI	4.0	Colon ca. SW-948	7.8
Melanoma* SK- MEL-5	4.2	Colon ca. SW480	32.3
Squamous cell carcinoma SCC-4	8.4	Colon ca.* (SW480 met) SW620	4.6
Testis Pool	1.1	Colon ca. HT29	30.6
Prostate ca.* (bone met) PC-3	24.8	Colon ca. HCT-116	5.8
Prostate Pool	0.8	Colon ca. CaCo-2	10.4
Placenta	1.3	Colon cancer tissue	10.0
Uterus Pool	0.4	Colon ca. SW1116	3.6
Ovarian ca. OVCAR-3	66.9	Colon ca. Colo-205	1.5
Ovarian ca. SK-OV- 3	36.3	Colon ca. SW-48	0.7
Ovarian ca. OVCAR-4	12.7	Colon Pool	1.3
Ovarian ca. OVCAR-5	44.4	Small Intestine Pool	1.0
Ovarian ca. IGROV- 1	27.7	Stomach Pool	1.2
Ovarian ca. OVCAR-8	14.9	Bone Marrow Pool	0.5
Ovary	1.9	Fetal Heart	1.0
Breast ca. MCF-7	9.7	Heart Pool	0.8
Breast ca. MDA- MB-231	31.2	Lymph Node Pool	2.0
Breast ca. BT 549	10.1	Fetal Skeletal Muscle	0.5
Breast ca. T47D	100.0	Skeletal Muscle Pool	0.5
Breast ca. MDA-N	4.2	Spleen Pool	0.7
Breast Pool	1.6	Thymus Pool	2.2
Trachea	2.6	CNS cancer (glio/astro) U87-MG	6.0
Lung	0.1	CNS cancer (glio/astro) U-118-MG	11.2
Fetal Lung	8.3	CNS cancer (neuro;met) SK-N-AS	0.9
Lung ca. NCI-N417	0.7	CNS cancer (astro) SF- 539	5.0
Lung ca. LX-1	11.0	CNS cancer (astro) SNB-75	32.3

Lung ca. NCI-H146	0.1	CNS cancer (glio) SNB-19	20.2
Lung ca. SHP-77	0.8	CNS cancer (glio) SF- 295	38.2
Lung ca. A549	10.4	Brain (Amygdala) Pool	1.3
Lung ca. NCI-H526	1.6	Brain (cerebellum)	1.0
Lung ca. NCI-H23	20.6	Brain (fetal)	2.8
Lung ca. NCI-H460	9.3	Brain (Hippocampus) Pool	0.9
Lung ca. HOP-62	23.0	Cerebral Cortex Pool	0.9
Lung ca. NCI-H522	2.3	Brain (Substantia nigra) Pool	1.7
Liver	0.6	Brain (Thalamus) Pool	1.6
Fetal Liver	1.4	Brain (whole)	1.1
Liver ca. HepG2	12.6	Spinal Cord Pool	1.4
Kidney Pool	2.5	Adrenal Gland	0.4
Fetal Kidney	4.6	Pituitary gland Pool	0.2
Renal ca. 786-0	39.5	Salivary Gland	1.3
Renal ca. A498	7.9	Thyroid (female)	3.7
Renal ca. ACHN	15.9	Pancreatic ca. CAPAN2	27.7
Renal ca. UO-31	38.7	Pancreas Pool	4.1

Table CMD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3974, Run 170739806	Tissue Name	Rel. Exp.(%) Ag3974, Run 170739806
Secondary Th1 act	1.2	HUVEC IL-1beta	18.9
Secondary Th2 act	8.0	HUVEC IFN gamma	16.7
Secondary Tr1 act	3.5	HUVEC TNF alpha + IFN gamma	34.9
Secondary Th1 rest	0.7	HUVEC TNF alpha + IL4	31.4
Secondary Th2 rest	0.2	HUVEC IL-11	13.9
Secondary Tr1 rest	1.2	Lung Microvascular EC none	100.0
Primary Th1 act	3.2	Lung Microvascular EC TNFalpha + IL-1beta	97.9
Primary Th2 act	2.0	Microvascular Dermal EC none	48.3
Primary Tr1 act	2.9	Microvascular Dermal EC TNFalpha + IL-1beta	47.0
Primary Th1 rest	0.4	Bronchial epithelium TNFalpha + IL1beta	90.1
Primary Th2 rest	0.2	Small airway epithelium	32.5

		none	
Primary Tr1 rest	0.3	Small airway epithelium TNFalpha + IL-1beta	93.3
CD45RA CD4 lymphocyte act	22.7	Coronary artery SMC rest	28.5
CD45RO CD4 lymphocyte act	5.5	Coronary artery SMC TNFalpha + IL-1beta	28.7
CD8 lymphocyte act	3.3	Astrocytes rest	55.1
Secondary CD8 lymphocyte rest	3.3	Astrocytes TNFalpha + IL-1beta	66.4
Secondary CD8 lymphocyte act	3.5	KU-812 (Basophil) rest	1.9
CD4 lymphocyte none	0.1	KU-812 (Basophil) PMA/ionomycin	2.8
2ry Th1/Th2/Tr1_anti- CD95 CH11	0.4	CCD1106 (Keratinocytes) none	82.4
LAK cells rest	6.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	72.7
LAK cells IL-2	1.7	Liver cirrhosis	14.4
LAK cells IL-2+IL-12	1.8	NCI-H292 none	54.0
LAK cells IL-2+IFN gamma	1.1	NCI-H292 IL-4	78.5
LAK cells IL-2+ IL-18	1.6	NCI-H292 IL-9	79.6
LAK cells PMA/ionomycin	4.6	NCI-H292 IL-13	59.9
NK Cells IL-2 rest	1.9	NCI-H292 IFN gamma	71.7
Two Way MLR 3 day	12.4	HPAEC none	21.3
Two Way MLR 5 day	5.3	HPAEC TNF alpha + IL-1 beta	45.4
Two Way MLR 7 day	4.0	Lung fibroblast none	29.3
PBMC rest	0.6	Lung fibroblast TNF alpha + IL-1 beta	87.7
PBMC PWM	4.9	Lung fibroblast IL-4	23.3
PBMC PHA-L	3.4	Lung fibroblast IL-9	30.4
Ramos (B cell) none	0.4	Lung fibroblast IL-13	36.6
Ramos (B cell) ionomycin	0.2	Lung fibroblast IFN gamma	29.7
B lymphocytes PWM	3.0	Dermal fibroblast CCD1070 rest	27.2
B lymphocytes CD40L and IL-4	3.7	Dermal fibroblast CCD1070 TNF alpha	20.6
EOL-1 dbcAMP	3.1	Dermal fibroblast CCD1070 IL-1 beta	22.4
EOL-1 dbcAMP PMA/ionomycin	8.0	Dermal fibroblast IFN gamma	10.3

Dendritic cells none	9.0	Dermal fibroblast IL-4	8.0
Dendritic cells LPS	32.8	Dermal Fibroblasts rest	6.3
Dendritic cells anti-CD40	8.8	Neutrophils TNFa+LPS	0.9
Monocytes rest	1.4	Neutrophils rest	1.0
Monocytes LPS	81.2	Colon	5.8
Macrophages rest	9.7	Lung	23.3
Macrophages LPS	43.8	Thymus	7.3
HUVEC none	12.6	Kidney	33.2
HUVEC starved	25.2		

CNS_neurodegeneration_v1.0 Summary: Ag3974 This panel does not show differential expression of the NOV98 gene in Alzheimer's disease. However, this expression profile confirms the presence of this gene in the brain. Please see Panel 1.3D for discussion of utility of this gene in the central nervous system.

General_screening_panel_v1.4 Summary: Ag3974 The expression of the NOV98 gene appears to be highest in a sample derived from a breast cancer cell line (T47D) (CT=22.5). In addition, there appears to be substantial expression in other samples derived from breast cancer cell lines, ovarian cancer cell lines, kidney cancer cell lines, lung cancer cell lines, colon cancer cell lines and brain cancer cell lines. Thus, the expression of this gene could be used to distinguish T47D cells from other samples in the panel. Moreover, therapeutic modulation of this gene, through the use of small molecule drugs, protein therapeutics, or antibodies could be of benefit in the treatment of breast, ovarian, kidney, lung, colon or brain cancer.

Among metabolic tissues, this gene has low-to-moderate levels of expression in adrenal, pituitary, adult and fetal heart, adult and fetal liver, adult and fetal skeletal muscle, and adipose. This gene product has high levels of expression (CT values = 27) in pancreas and thyroid. Thus, this gene product may be important for the pathogenesis, diagnosis, and/or treatment of metabolic and endocrine diseases, including obesity, Types 1 and 2 diabetes and thyroidopathies. It has recently been reported that an agrin minigene rescued dystrophic symptoms in a mouse model of muscular dystrophy. Therefore, this gene product may also be used as a treatment or cure for congenital muscular dystrophies.

This gene is also expressed at moderate to high levels in all regions of the CNS. This molecule is a homolog of agrin, which has been implicated in the formation of senile plaques in Alzheimer's disease and in the acetylcholine synapse/neuromuscular junction. This gene is

therefore an excellent drug target in AD or in any disease involving the neuromuscular junction or the acetylcholine system.

References:

- 5 Moll J, Barzaghi P, Lin S, Bezakova G, Lochmuller H, Engvall E, Muller U, Ruegg MA. An agrin minigene rescues dystrophic symptoms in a mouse model for congenital muscular dystrophy. *Nature*. 2001 Sep 20;413(6853):302-7.

10 Congenital muscular dystrophy is a heterogeneous and severe, progressive muscle-wasting disease that frequently leads to death in early childhood. Most cases of congenital muscular dystrophy are caused by mutations in LAMA2, the gene encoding the alpha2 chain of the main laminin isoforms expressed by muscle fibres. Muscle fibre deterioration in this disease is thought to be caused by the failure to form the primary laminin scaffold, which is necessary for basement membrane structure, and the missing interaction between muscle basement membrane and the dystrophin-glycoprotein complex (DGC) or the integrins. With the aim to restore muscle function in a mouse model for this disease, we have designed a minigene of agrin, a protein known for its role in the formation of the neuromuscular junction. Here we show that this mini-agrin-which binds to basement membrane and to alpha-dystroglycan, a member of the DGC-amends muscle pathology by a mechanism that includes agrin-mediated stabilization of alpha-dystroglycan and the laminin alpha5 chain. Our data provides in vivo evidence that a non-homologous protein in combination with rational protein design can be used to devise therapeutic tools that may restore muscle function in human muscular dystrophies.

PMID: 11565031

- 25 Liyanage Y, Hoch W, Beeson D, Vincent A. The agrin/muscle-specific kinase pathway: New targets for autoimmune and genetic disorders at the neuromuscular junction. *Muscle Nerve* 2002 Jan;25(1):4-16

30 The increasing understanding of the structural complexity of the neuromuscular junction (NMJ), and the processes that are important in its development, suggests many possible new disease targets. Here, we summarize briefly the genetic and autoimmune disorders that affect neuromuscular transmission, and the identified targets, including new evidence that antibodies to muscle-specific receptor tyrosine kinase (MuSK) are involved in the pathogenesis of acetylcholine receptor (AChR) antibody-negative myasthenia gravis. We then review the development of the NMJ, focusing on the important roles of nerve-derived agrin and MuSK in clustering of AChRs and other essential components of the NMJ.

van Horssen J, Otte-Holler I, David G, Maat-Schieman ML, van den Heuvel LP, Wesseling P, de Waal RM, Verbeek MM. Heparan sulfate proteoglycan expression in cerebrovascular amyloid beta deposits in Alzheimer's disease and hereditary cerebral hemorrhage with amyloidosis (Dutch) brains. *Acta Neuropathol (Berl)* 2001 Dec;102(6):604-14

Cerebrovascular deposition of amyloid beta protein (A beta) is a characteristic lesion of Alzheimer's disease (AD) and hereditary cerebral hemorrhage with amyloidosis of the Dutch type (HCHWA-D). Besides A beta, several other proteins and proteoglycans accumulate in cerebral amyloid angiopathy (CAA). We have now analyzed the expression of the heparan sulfate proteoglycan (HSPG) subtypes agrin, perlecan, glypican-1, syndecans 1-3 and HS glycosaminoglycan (GAG) side chains in CAA in brains of patients with AD and HCHWA-D. Hereto, specific well-characterized antibodies directed against the core protein of these HSPGs and against the GAG side chains were used for immunostaining. Glypican-1 was abundantly expressed in CAA both in AD and HCHWA-D brains, whereas perlecan and syndecans-1 and -3 were absent in both. Colocalization of agrin with vascular A beta was clearly observed in CAA in HCHWA-D brains, but only in a minority of the AD cases. Conversely, syndecan-2 was frequently associated with vascular A beta in AD, but did not colocalize with vascular A beta deposits in HCHWA-D. The three different syndecans, agrin, glypican-1 and HS GAG, but not perlecan, were associated with the majority of senile plaques (SPs) in all brains. Our results suggest a role for agrin in the formation of SPs and of CAA in HCHWA-D, but not in the pathogenesis of CAA in AD. Both syndecan-2 and glypican, but not perlecan, may be involved in the formation of CAA. We conclude that specific HSPG species may be involved in the pathogenesis of CAA in both AD and HCHWA-D, and that the pathogenesis of CAA and SPs may differ with regard to the involvement of HSPG species.

Panel 4.1D Summary: Ag3974 The NOV98 gene is expressed at moderate levels (CT=29-32) in a wide range of cell types of significance in the immune response in health and disease. Therefore, targeting of this gene product with a small molecule drug or antibody therapeutic may modulate the functions of cells of the immune system as well as resident tissue cells and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as COPD, emphysema, asthma, allergies, inflammatory bowel disease, lupus erythematosus, and arthritis, including osteoarthritis and rheumatoid arthritis. Based on its homology to agrin, this gene product may also be beneficial to the treatment of multiple sclerosis as suggested by the referene below.

References:

Liyanage Y, Hoch W, Beeson D, Vincent A. The agrin/muscle-specific kinase pathway: New targets for autoimmune and genetic disorders at the neuromuscular junction. *Muscle Nerve* 2002 Jan;25(1):4-16

The increasing understanding of the structural complexity of the neuromuscular junction (NMJ), and the processes that are important in its development, suggests many possible new disease targets. Here, we summarize briefly the genetic and autoimmune disorders that affect neuromuscular transmission, and the identified targets, including new evidence that antibodies to muscle-specific receptor tyrosine kinase (MuSK) are involved in the pathogenesis of acetylcholine receptor (AChR) antibody-negative myasthenia gravis. We then review the development of the NMJ, focusing on the important roles of nerve-derived agrin and MuSK in clustering of AChRs and other essential components of the NMJ.

Example 3. SNP analysis of NOVX clones

SeqCalling™ Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Variant sequences are included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a

nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

Method of novel SNP Identification: SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2; Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

Method of novel SNP Confirmation: SNPs are confirmed employing a validated method known as Pyrosequencing (Pyrosequencing, Westborough, MA). Detailed protocols for Pyrosequencing can be found in: Alderborn et al. Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. (2000). *Genome Research*. 10, Issue 8, August. 1249-1265. In brief, Pyrosequencing is a real time primer extension process of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminescent assay of pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase-mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes

proportional to the number of added bases, up to about four bases. To allow processivity of the method dNTP excess is degraded by apyrase, which is also present in the starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels. The DNA and protein sequences for the novel single nucleotide polymorphic variants are reported. Variants are reported individually but any combination of all or a select subset of variants are also included. In addition, the positions of the variant bases and the variant amino acid residues are underlined.

Results

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV1a SNP data:

NOV1a has two SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:1 and 2, respectively. The nucleotide sequence of the NOV1a variant differs as shown in Table 101.

Table 101. cSNP and Coding Variants for NOV1a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
393	T	C	130	S->P
420	T	C	139	W->R
431	T	C	142	No change
501	C	T	166	L->F
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
420	T	C	20	0.100

NOV1b SNP data:

NOV1b has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:3 and 4, respectively. The nucleotide sequence of the NOV1b variant differs as shown in Table 102.

Table 102. cSNP and Coding Variants for NOV1b				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
393	T	C	130	S->P
420	T	C	139	W->R

NOV3a SNP data:

NOV3a has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:11 and 12, respectively. The nucleotide sequence of the NOV3a variant differs as shown in Table 103.

Table 103. cSNP and Coding Variants for NOV3a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
212	T	C	54	C->R
439	A	G	149	No Change

NOV4a SNP data:

NOV4a has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:17 and 18, respectively. The nucleotide sequence of the NOV4a variant differs as shown in Table 104.

Table 104. cSNP and Coding Variants for NOV4a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
229	G	A	73	No change
390	G	A	127	W->End
631	G	C	207	Q->H

NOV5a SNP data:

NOV5a has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:23 and 24, respectively. The nucleotide sequence of the NOV5a variant differs as shown in Table 105.

Table 105. cSNP and Coding Variants for NOV5a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
79	C	T	19	A->V
204	T	C	61	No change
658	A	T	212	Q->L
884	C	G	287	No change
1149	G	T	376	D->Y

NOV6 SNP data:

NOV6 has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:33 and 34, respectively. The nucleotide sequence of the NOV6 variant differs as shown in Table 106.

Table 106. cSNP and Coding Variants for NOV6				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
543	T	C	156	V->A
549	T	C	158	I->T
660	C	T	195	A->V
734	A	G	220	T->A
782	G	A	236	A->T

5

NOV7a SNP data:

NOV7a has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:35 and 36, respectively. The nucleotide sequence of the NOV7a variant differs as shown in Table 107.

10

Table 107. cSNP and Coding Variants for NOV7a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid Position	Amino Acid Change
168	T	C	24	V->G
459	C	T	121	A->V
815	T	C	240	S->P
896	A	G	N/A	No change
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
428	G	A	8	0.250

NOV7c SNP data:

NOV7c has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:39 and 40, respectively. The nucleotide sequence of the NOV7c variant differs as shown in Table 108.

15

Table 108. cSNP and Coding Variants for NOV7c				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
383	C	T	5	0.400

NOV7d SNP data:

NOV7d has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:41 and 42, respectively. The nucleotide sequence of the NOV7d variant differs as shown in Table 109.

Table 109. cSNP and Coding Variants for NOV7d				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid Position	Amino Acid Change
260	G	A	86	I->E

NOV7e SNP data:

NOV7e has four SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:323 and 324, respectively. The nucleotide sequence of the NOV7d variant differs as shown in Table 110.

Table 110. cSNP and Coding Variants for NOV7e				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid Position	Amino Acid Change
304	A	G	102	T->A

NOV9a SNP data:

NOV9a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:45 and 46, respectively. The nucleotide sequence of the NOV9a variant differs as shown in Table 111.

Table 111. cSNP and Coding Variants for NOV9a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
264	G	A	N/A	No change
391	C	A	8	P->T
438	G	A	23	No change
550	T	G	61	F->V
672	C	T	101	No change
1286	T	C	306	L->S
1338	G	A	323	No change

NOV10 SNP data:

NOV10 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:49 and 50, respectively. The nucleotide sequence of the NOV10 variant differs as shown in Table 112.

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
335	G	A	10	0.400

NOV13b SNP data:

NOV13b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:57 and 58, respectively. The nucleotide sequence of the NOV13b variant differs as shown in Table 113.

5

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
362	G	A	11	0.455

NOV15b SNP data:

NOV15b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:63 and 64, respectively. The nucleotide sequence of the NOV15b variant differs as shown in Table 114.

10

Table 114. cSNP and Coding Variants for NOV15b				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
388	T	G	17	0.294

NOV16b SNP data:

15 NOV16b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:67 and 68, respectively. The nucleotide sequence of the NOV16b variant differs as shown in Table 115.

15

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
463	A	T	16	0.125
465	C	T	16	0.125
535	T	C	15	0.133
735	C	T	12	0.167

814	T	G	11	0.182
-----	---	---	----	-------

NOV21a SNP data:

NOV21a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:87 and 88, respectively. The nucleotide sequence of the NOV21a variant differs as shown in Table 116.

Table 116. cSNP and Coding Variants for NOV21a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
121	T	C	41	W->R
170	T	C	57	V->A
364	C	T	122	No change
415	G	A	139	E->K
535	G	A	182	R->H
630	G	A	210	No change

NOV21b SNP data:

NOV21b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:89 and 90, respectively. The nucleotide sequence of the NOV21b variant differs as shown in Table 117.

Table 117. cSNP and Coding Variants for NOV21b				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
485	G	A	65	0.246
616	G	A	44	0.136
714	G	A	36	0.083

NOV22a SNP data:

NOV22a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:91 and 92, respectively. The nucleotide sequence of the NOV22a variant differs as shown in Table 118.

Table 118. cSNP and Coding Variants for NOV22a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
669	A	G	223	No change
725	C	T	242	T->M

NOV22c SNP data:

NOV22c has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:95 and 96, respectively. The nucleotide sequence of the NOV22c variant differs as shown in Table 119.

5

Table 119. cSNP and Coding Variants for NOV22c				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
85	A	G	30	0.067
288	T	C	31	0.065
484	A	G	37	0.054
540	C	T	29	0.241

NOV24a SNP data:

NOV24a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:103 and 104, respectively. The nucleotide sequence of the NOV24a variant differs as shown in Table 120.

10

Table 120. cSNP and Coding Variants for NOV24a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
539	C	T	511	No change

NOV24b SNP data:

NOV24b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:105 and 106, respectively. The nucleotide sequence of the NOV24b variant differs as shown in Table 121.

15

Table 121. cSNP and Coding Variants for NOV24b				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
437	A	G	143	N->S
664	T	G	219	F->V
1150	G	T	381	A->S
1210	G	T	401	E->End
1770	C	T	587	No change
2011	A	G	N/A	No change
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
329	C	T	11	0.364

491	A	C	13	0.154
-----	---	---	----	-------

NOV25 SNP data:

NOV25 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:107 and 108, respectively. The nucleotide sequence of the NOV25 variant differs as shown in Table 122.

Table 122. cSNP and Coding Variants for NOV25				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
221	G	A	54	No change
462	C	T	135	L->F

NOV26a SNP data:

NOV26a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:109 and 110, respectively. The nucleotide sequence of the NOV26a variant differs as shown in Table 123.

Table 123. cSNP and Coding Variants for NOV26a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
67	A	G	1	M->V
98	C	T	11	P->L
128	A	G	21	E->G
176	A	G	37	A->T
233	A	G	56	Q->R
243	T	C	59	No change
252	A	G	62	No change
260	A	G	65	D->G
296	A	G	77	K->R
316	A	G	84	N->D
369	G	A	101	M->I
395	A	G	110	Q->R
465	G	A	N/A	No change

NOV26b SNP data:

NOV26b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:111 and 112, respectively. The nucleotide sequence of the NOV26b variant differs as shown in Table 124.

Table 124. cSNP and Coding Variants for NOV26b

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
133	A	G	41	0.049
268	A	G	41	0.049
324	A	G	41	0.049
372	A	G	41	0.049
376	A	*	41	0.049
456	T	C	40	0.050
488	A	G	32	0.344

NOV27a SNP data:

NOV27a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:113 and 114, respectively. The nucleotide sequence of the NOV27a variant differs as shown in Table 125.

Table 125. cSNP and Coding Variants for NOV27a

NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
158	A	T	36	No change
491	C	T	147	No change
562	T	C	171	L->P
858	C	T	270	No change
1750	C	T	567	P->L

NOV27b SNP data:

NOV27b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:115 and 116, respectively. The nucleotide sequence of the NOV27b variant differs as shown in Table 126.

Table 126. cSNP and Coding Variants for NOV27b

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
131	G	A	43	0.233
1072	G	A	14	0.214
1368	A	G	24	0.125
1439	G	A	42	0.071
1733	G	A	43	0.047
1772	T	A	43	0.442
1787	G	A	42	0.286

NOV29c SNP data:

NOV29c has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:125 and 126, respectively. The nucleotide sequence of the NOV29c variant differs as shown in Table 127.

5

Table 127. cSNP and Coding Variants for NOV29c				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
760	A	G	254	T->A
923	T	C	308	G->D

NOV30 SNP data:

NOV30 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:127 and 128, respectively. The nucleotide sequence of the NOV30 variant differs as shown in Table 128.

10

Table 128. cSNP and Coding Variants for NOV30				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
103	A	G	28	I->V
207	T	C	62	No change
225	C	T	68	No change
233	A	G	71	D->G
267	T	C	82	No change
318	A	G	99	No change
392	T	C	124	L->P
431	T	C	137	M->T
464	A	G	148	E->G
479	T	A	153	V->E

NOV33 SNP data:

NOV33 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:133 and 134, respectively. The nucleotide sequence of the NOV33 variant differs as shown in Table 129.

15

Table 129. cSNP and Coding Variants for NOV33				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
5097	T	C	1699	No change
6012	C	T	2004	No change

NOV36a SNP data:

NOV36a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:139 and 140, respectively. The nucleotide sequence of the NOV36a variant differs as shown in Table 130.

5

Table 130. cSNP and Coding Variants for NOV36a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
351	T	C	102	No change
737	A	G	231	D->G

NOV38 SNP data:

NOV38 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:145 and 146, respectively. The nucleotide sequence of the NOV38 variant differs as shown in Table 131.

10

Table 131. cSNP and Coding Variants for NOV38				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
566	C	T	188	H->R
658	A	G	219	H->R
844	G	A	281	C->Y
892	C	T	297	A->V
910	T	C	303	V->A
1009	G	A	336	S->N
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
95	T	C	14	N/A

NOV39a SNP data:

NOV39a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:147 and 148, respectively. The nucleotide sequence of the NOV39a variant differs as shown in Table 132.

15

Table 132. cSNP and Coding Variants for NOV39a				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
1095	T	C	11	N/A

NOV39b SNP data:

NOV39b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:149 and 150, respectively. The nucleotide sequence of the NOV39b variant differs as shown in Table 133.

Table 133. cSNP and Coding Variants for NOV39b				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
933	C	T	9	0.222

NOV42c SNP data:

NOV42c has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:161 and 162, respectively. The nucleotide sequence of the NOV42c variant differs as shown in Table 134.

Table 134. cSNP and Coding Variants for NOV42c				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
330	G	A	103	R->Q
783	A	G	254	D->G
903	G	A	294	R->H
1389	A	T	456	E->V
1389	A	G	456	E->G
1394	G	A	458	A->T
1642	C	T	540	No change
1656	T	C	545	V->A
1658	G	A	546	A->T

NOV43 SNP data:

NOV43 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:165 and 166, respectively. The nucleotide sequence of the NOV43 variant differs as shown in Table 135.

Table 135. cSNP and Coding Variants for NOV43				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
378	G	A	121	No change
496	G	A	162	R->Q

NOV46b SNP data:

NOV46b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:173 and 174, respectively. The nucleotide sequence of the NOV46b variant differs as shown in Table 136.

Table 136. cSNP and Coding Variants for NOV46b				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
500	C	T	163	No change
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
486	C	T	16	0.125

NOV48a SNP data:

NOV48a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:181 and 182, respectively. The nucleotide sequence of the NOV48a variant differs as shown in Table 137.

Table 137. cSNP and Coding Variants for NOV48a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
370	A	G	1	M->V
436	C	T	23	L->F
539	A	G	57	D->G
650	A	G	94	E->G
1012	C	T	215	Q->End
1922	A	G	518	K->R
2057	A	G	563	Q->R
2066	C	T	566	A->V
2198	C	T	610	P->L
2618	A	G	750	D->G
2656	G	A	N/A	No change

NOV50b SNP data:

NOV50b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:189 and 190, respectively. The nucleotide sequence of the NOV50b variant differs as shown in Table 138.

Table 138. cSNP and Coding Variants for NOV50b				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
797	A	G	265	Q->R

NOV52 SNP data:

NOV52 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:193 and 194, respectively. The nucleotide sequence of the NOV52 variant differs as shown in Table 139.

Table 139. cSNP and Coding Variants for NOV52				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
318	T	C	48	No change
351	C	T	59	R->C
1961	G	T	595	M->I
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
70	C	T	54	0.056

NOV56a SNP data:

NOV56a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:203 and 204, respectively. The nucleotide sequence of the NOV56a variant differs as shown in Table 140.

Table 140. cSNP and Coding Variants for NOV56a				
NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
118	A	C	39	No change
435	C	T	135	P->L
439	T	C	146	No change
473	A	G	158	T->A
588	T	C	196	V->A
596	G	A	199	G->R
614	A	G	205	M->V
631	T	C	210	No change
637	A	G	212	No change
642	T	C	214	M->T
732	G	T	244	W->L
902	A	T	301	M->L

NOV57 SNP data:

NOV57 has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:207 and 208, respectively. The nucleotide sequence of the NOV57 variant differs as shown in Table 141.

Table 141. cSNP and Coding Variants for NOV57

NT Position of cSNP	Wild Type NT	Variant NT	Amino Acid position	Amino Acid Change
939	T	A	N/A	No change

NOV58b SNP data:

NOV58b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:211 and 212, respectively. The nucleotide sequence of the NOV58b variant differs as shown in Table 142.

Table 142. cSNP and Coding Variants for NOV58b

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
88	T	C	11	0.273
377	A	G	18	0.111
500	T	C	18	0.111
509	A	G	18	0.111
570	T	C	17	0.118
647	C	T	9	0.222

NOV60a SNP data:

NOV60a has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:215 and 216, respectively. The nucleotide sequence of the NOV60a variant differs as shown in Table 143.

Table 143. cSNP and Coding Variants for NOV60a

NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
341	T	C	52	
401	G	C	79	
411	T	C	79	
444	C	T	79	
644	C	T	59	
653	T	A	59	
670	T	C	59	
707	T	C	33	

NOV60b SNP data:

NOV60b has one SNP variant, whose variant position for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:217 and 218, respectively. The nucleotide sequence of the NOV60b variant differs as shown in Table 144.

Table 144. cSNP and Coding Variants for NOV60b				
NT Position of cSNP	Wild Type NT	Variant NT	Depth	Putative Allele Frequency
162	T	G	54	0.259
192	A	G	54	0.056
229	A	G	81	0.025
246	A	T	80	0.025
255	G	A	79	0.038
263	A	G	77	0.039
342	G	A	103	0.019
389	A	G	105	0.019

Example 4. In-frame Cloning

NOV7c

For NOV7c the cDNA coding for the DOMAIN of NOV7c from residues 1 to 230 was targeted for “in-frame” cloning by PCR. The PCR template was based on the previously identified plasmid, when available, or on human cDNA(s).

Table 145. Oligonucleotide primers used to clone the target cDNA sequence:

Primers	Sequences
F1	5' - AGATCTCCACCC ATGCAACTTCAGGACCTGGAAGTGC -3' (SEQ ID NO:1382)
R1	5' - CTCGAG TCCACTTACAATTTCCTGCTGATTTC -3' (SEQ ID NO:1385)
SF1	5' - TCCTCCTGGAGAAAGCTCAGAATCTGTTTT -3' (SEQ ID NO:1387)
SF2	5' - CTCCAGATTGGAAGTTCTGAGGAA -3' (SEQ ID NO:1388)
SR1	5' - ATTTCTCCAAGTCCCAGGCC -3' (SEQ ID NO:1389)
SR2	5' - GAGCCTGTCTAGAAAGGAGCTGTG -3' (SEQ ID NO:1390)

For downstream cloning purposes, the forward primer includes an in-frame Hind III restriction site and the reverse primer contains an in-frame Xho I restriction site.

Two parallel PCR reactions were set up using a total of 0.5-1.0 ng human pooled cDNAs as template for each reaction. The pool is composed of 5 micrograms of each of the following human tissue cDNAs: adrenal gland, whole brain, amygdala, cerebellum, thalamus, bone marrow, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, liver, lymphoma, Burkitt's Raji cell line, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small Intestine, spleen, stomach, thyroid, trachea, uterus.

When the tissue of expression is known and available, the second PCR was performed using the above primers and 0.5ng-1.0 ng of one of the following human tissue cDNAs:

skeleton muscle, testis, mammary gland, adrenal gland, ovary, colon, normal cerebellum, normal adipose, normal skin, bone marrow, brain amygdala, brain hippocampus, brain substantia nigra, brain thalamus, thyroid, fetal lung, fetal liver, fetal brain, kidney, heart, spleen, uterus, pituitary gland, lymph node, salivary gland, small intestine, prostate, placenta, spinal cord, peripheral blood, trachea, stomach, pancreas, hypothalamus.

The reaction mixtures contained 2 microliters of each of the primers (original concentration: 5 pmol/ul), 1 microliter of 10mM dNTP (Clontech Laboratories, Palo Alto CA) and 1 microliter of 50xAdvantage-HF 2 polymerase (Clontech Laboratories) in 50 microliter-reaction volume. The following reaction conditions were used:

PCR condition 1:

- a) 96°C 3 minutes
- b) 96°C 30 seconds denaturation
- c) 60°C 30 seconds, primer annealing
- d) 72°C 6 minutes extension

Repeat steps b-d 15 times

- e) 96°C 15 seconds denaturation
- f) 60°C 30 seconds, primer annealing
- g) 72°C 6 minutes extension

Repeat steps e-g 29 times

- e) 72°C 10 minutes final extension

PCR condition 2:

- a) 96°C 3 minutes
- b) 96°C 15 seconds denaturation
- c) 76°C 30 seconds, primer annealing, reducing the temperature by 1 °C per cycle
- d) 72°C 4 minutes extension

Repeat steps b-d 34 times

- e) 72°C 10 minutes final extension

Example 5: SAGE Analysis

Hs 181638 : ESTs, Weakly similar to SSR1_HUMAN SOMATOSTATIN RECEPTOR TYPE 1 [H.sapiens]

SAGE library data and reliable tag summary

Reliable tags found in SAGE libraries:

<u>Library name</u>	<u>Tags per million</u>	<u>Tag counts</u>	<u>Total tags</u>
<u>SAGE Chen LNCaP</u>	16	1	62267
<u>SAGE Chen Normal Pr</u>	30	2	66193
<u>SAGE Chen Tumor Pr</u>	14	1	68384
<u>SAGE CAPAN1</u>	26	1	37926
<u>SAGE Panc1</u>	80	2	24879
<u>SAGE Duke H54 EGFRvIII</u>	34	2	57164
<u>SAGE CPDR LNCaP-T</u>	22	1	44122
<u>SAGE 293-IND</u>	40	1	24481
<u>SAGE PR317 normal prostate</u>	16	1	59419
<u>SAGE PR317 prostate tumor</u>	15	1	65109
<u>SAGE BB542 whitmatter</u>	10	1	94806
<u>SAGE Panc 96-6252</u>	27	1	35745
<u>SAGE SciencePark MCF7</u>	16	1	61079
<u>Control 0h</u>			
<u>SAGE SciencePark MCF7</u>	16	1	60435
<u>estradiol 10h</u>			
<u>SAGE Duke H566</u>	15	1	65728
<u>SAGE OVT-6</u>	23	1	42336
<u>SAGE mammary epithelium</u>	20	1	49167
<u>SAGE ML10-10</u>	35	2	56943
<u>SAGE Duke H1043</u>	13	1	76673

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is

5 contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the

10 embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.